

ABSTRACT

Title of thesis: DOUBLY PENALIZED LOGISTIC REGRESSION FOR
GENOMEWIDE ASSOCIATION STUDIES WITH
LINEARLY STRUCTURED GENETIC NETWORKS

Xia Li, Master of Public Health, 2012

Directed by: Assistant Professor Tongtong Wu
Department of Epidemiology and Biostatistics

This research aims to integrate linear structures of genetic networks into genomewide analysis studies (GWAS). Lasso penalized logistic regression is ideally suited for continuous model selection in case-control disease gene mapping, especially when the number of predictor variables far exceeds the number of observations. But it fails to consider the structure of genetic networks. Imposing an additional weighted fused lasso can further remove irrelevant predictors. Nesterov's method is employed to handle the high dimensionality and complexity of genetic data. It also resolves the non-differentiability problem of the lasso and fused lasso penalties. In simulation studies, this proposed method shows advantages in some cases compared with lasso and fused lasso. We apply this method to the coeliac data on chromosome 8.

DOUBLY PENALIZED LOGISTIC REGRESSION FOR
GENOMEWIDE ASSOCIATION STUDIES WITH
LINEARLY STRUCTURED GENETIC NETWORKS

By

Xia Li

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Public Health
2012

Advisory Committee:

Assistant Professor Tongtong Wu, Chair

Assistant Professor Guangyu Zhang

Associate Professor Paul J. Smith

© Copyright by
Xia Li
2012

Table of Contents

List of Tables	iii
List of Figures	iv
List of Abbreviations	v
1 Introduction	1
2 Background	5
2.1 The Lasso Penalty	5
2.2 The Lasso Estimate	7
2.3 Fused Lasso and Linear Structure of SNPs	9
2.4 Selection of Tuning Parameters	10
3 Method	12
3.1 Nesterov's Method	13
3.2 Fused Lasso Signal Approximator	14
3.3 Nested Procedure of Nesterov's Method and FLSA	16
3.4 Selection of Tuning Variables	17
3.4.1 Validation Method	17
3.4.2 BIC Method and Its Extensions	18
4 Analysis of Simulated Data	19
4.1 Input Setting of Simulated Data	19
4.2 Simulation Study Results	20
5 Real Data Analysis	23
5.1 Data Description	24
5.2 Results of Real Data Analysis	24
6 Conclusion	28
Appendix	43
References	43

List of Tables

1	Simulation results for $\rho = 0$ based on validation method	30
2	Simulation results for $\rho = 0$ based on BIC	31
3	Simulation results for $\rho = 0$ based on BICC	32
4	Simulation results for $\rho = 0$ based on EBIC	33
5	Simulation results for $\rho = 0.5$ based on validation method	34
6	Simulation results for $\rho = 0.5$ based on BIC	35
7	Simulation results for $\rho = 0.5$ based on BICC	36
8	Simulation results for $\rho = 0.5$ based on EBIC	37
9	Simulation results for $\rho = 0.9$ based on validation method	38
10	Simulation results for $\rho = 0.9$ based on BIC	39
11	Simulation results for $\rho = 0.9$ based on BICC	40
12	Simulation results for $\rho = 0.9$ based on EBIC	41
13	Selected predictors in chromosome 8 for coeliac disease	42

List of Figures

- 2.1 Estimate picture for (a) the lasso and (b) ridge regression (Tibshirani 1996) 7

List of Abbreviations

BIC	Bayesian Information Criterion
EBIC	Extended Bayesian Information Criterion
FLSA	Fused Lasso Signal Approximator
GWAS	Genome Wide Association Study
Lasso	Least Absolute Shrinkage and Selection Operator
OLS	Ordinary Least Square
SNP	Single Nucleotide Polymorphisms

Chapter 1

Introduction

Both environmental and genetic factors have roles in the development of any disease. As a risk behavior, tobacco smoking has been found to be associated with many diseases including heart attacks, strokes, emphysema, lung cancer and pancreatic cancer ([Sasco et al. 2004](#)). About 443,000 U.S. deaths are attributable each year to tobacco smoking ([CDC 2011](#)). Vegetable and fruit consumption is protective against coronary heart disease ([Dauchet et al. 2006](#)), diabetes ([Ford and Mokdad 2001](#)), stroke ([He et al. 2006](#)), and cancer ([Steinmetz and Potter 1996](#)).

Unlike those factors that can change, genetic factors are usually unchangeable and can also alter risks for disease incidences. Substantial efforts have been made to identify all common genetic variations in humans, including single nucleotide polymorphisms (SNPs), deletions and insertions ([Brookes 1999](#)). For example, researchers have found that the genes BRCA1 and BRCA2 promote breast ([Miki et al. 1994](#); [Ford et al. 1998](#)), ovarian ([Miki et al. 1994](#)) and pancreatic cancers ([Venkitaraman 2002](#)). On occasion, some genetic factors may be beneficial in a given environment. For example, [Sullivan et al. \(2001\)](#) found that CCR5 protect against HIV infection.

A genome-wide association study (GWAS) is an approach that identifies common genetic factors associated with a particular disease. GWAS is usually useful in finding genetic variations that contribute to common and complex diseases. SNP is a single nucleotide change in the DNA sequence. Such common genetic variation occurs in both coding and non-coding regions of genome when a single nucleotide, such as an A, replaces

one of the other three nucleotide, C, G, or T (Brookes 1999; Vignal et al. 2002; NIH 2011). For a base position to be considered as a SNP, the least frequent allele should have a frequency of 1% or greater among the specific population. Mutations, on the other hand, are changes in the DNA or RNA sequence. Types of mutation include substitution of a single nucleotide (no 1% restriction), insertation of extra base pairs, deletion, and frame shift. Compared with SNP, mutation is relatively rare and new in population but may directly cause disorder. The human genome has about 3 billion base pairs of nucleotides. On average SNPs occur once in every 300 nucleotides, which means there are about 10 million SNPs in the human genome (NIH 2011). SNPs are believed to alter the risk for developing particular diseases. It is, however, very unlikely that individual SNP plays an important role in the development of complex diseases. Instead, high-order interactions of SNPs are thought to explain the differences between low and high-risk population groups.

A simple and plausible way to examine the association between SNPs and disease is to apply logistic regression to identify potential strong predictors in case-control studies. One issue with this, however, is that the number of gene expression profiles is often overwhelmingly large, in the tens of thousands, far exceeding the number of subjects. This imposes problems in terms of both theory and computation. The usual multivariate regression methods break down because matrix inversion is not feasible. A statistical computational method which can achieve feature selection and estimation is needed.

The lasso penalty has been developed and used as a powerful tool yielding regression estimates with many coefficients set to zero (Claerbout and Muir 1973; Taylor et al. 1979; Santosa and Symes 1986; Tibshirani 1996; Chen et al. 1999). In signal processing, the lasso is also known as *basis pursuit* (Chen et al. 1999). Consider an ordinary regression data set $\{y, x_1, \dots, x_n\}$ where $y = (y_1, \dots, y_n)^t$ is the response vector and $x_i = (x_{i1}, \dots, x_{ip})^t$ is

the predictor vector for the i^{th} subject. Let $\theta = (\mu, \beta_1, \dots, \beta_p)^t$ be the parameter vector; μ is the intercept. The lasso penalty is the L_1 norm of coefficients, $\sum_{j=1}^p |\beta_j|$, which is constrained to be smaller than a given positive value. Equivalently, in ordinary linear regression the lasso estimate is the solution to

$$\min_{\theta} \left\{ \sum_{i=1}^n (y_i - \mu - x_i^t \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}, \quad (1.1)$$

where λ is a tuning constant chosen by the statistician. It works very well as a variable selection tool, especially in cases with a large number of predictors. Unlike ridge regression, a similar variable selection method, lasso can determine an easy-to-interpret model by forcing some coefficients strictly to 0. Several authors have explored lasso penalized ordinary regression. [Fu \(1998\)](#) gave the Shooting Algorithm for the lasso by studying the structure of the bridge estimators; [Knight and Fu \(2000\)](#) proved some asymptotic properties for lasso-type estimators. [Fan and Li \(2001\)](#) studied the penalized likelihood methods in linear regression, of which the lasso is a special case. [Daubechies et al. \(2004\)](#) proposed an iterative shrinkage/thresholding (IST) algorithm which also could deal with optimizing $f(\theta)$. [Osborne et al. \(2000\)](#) developed a new algorithm based on the consideration of primal and dual problems. Another possible method is gradient lasso, which is computationally more stable than quadratic program based algorithms ([Kim and Kim 2004](#); [Kim et al. 2008](#)). The papers of [Friedman et al. \(2007\)](#) and [Wu and Lange \(2008\)](#) chose coordinate descent to solve the lasso penalized ordinary regression. Some possible extensions of the lasso penalty to generalized linear models haven been discussed ([Fu 1998](#); [Park and Hastie 2007, 2008](#)). Competing algorithms for lasso penalized regression include non-negative quadratic programming ([Sha et al. 2007](#)), quadratic approximations ([Lee et al. 2006](#)), interior point methods ([Koh et al. 2007](#)) and coordinate descent methods

(Wu and Lange 2008; Friedman et al. 2007). Friedman et al. (2010) concluded that coordinate descent performs the best. Wu et al. (2009) further explored the cyclic coordinate ascent in logistic regression, where the penalized likelihood can be quickly maximized for a given tuning constant.

$$\sum_{i=1}^n (y_i - \mu - x_i^t \beta)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=2}^p |\beta_j - \beta_{j-1}| \quad (1.2)$$

The fused lasso, first proposed by Tibshirani et al. (2005), is an improved instrument for variable selection with ordered features. It minimizes the criterion (1.2). However, the coordinate descent and its extensions may not converge to the desired solution when applied to the fused lasso problem. Because the fused lasso problem is not separable and coordinate descent works well in separable problems. Friedman et al. (2007) modified the coordinate-wise descent procedure into fused lasso signal approximator (FLSA) for solving a special case of (1.2). Rinaldo (2009) modified the fused lasso estimator in a signal estimation problem. Another adaptive form, weighted fusion, was proposed allowing the selection of more than the number of observations (Daye and Jeng 2009). Liu et al. (2010) solved the fused lasso problem, combining Nesterov’s optimal first-order method (Nesterov 1983, 2003) and procedure developed by Friedman et al. (2007).

In this thesis, we will consider the linear structure of SNPs. Variation in SNPs only explains a small fraction of disease; most of the detectable odds ratios are between 1.1 and 1.3 (Goldstein 2009). It is likely that there are many more common variants that have not been detected by GWAS because they alter the risk by smaller values, perhaps as low as 1% (Cantor et al. 2010). In addition, incorporation of linear structures aim to increase the chance of detecting the SNPs with weak association. We will mimic the procedure nested between Nesterov’s strategy and FLSA, used in paper of Liu et al. (2010)

Chapter 2

Background

In GWAS, we compare regions of the genome between cohorts and try to tease out the irrelevant and redundant SNPs. We are unable to interpret the model if all the SNPs are considered. On the other hand, it's not feasible to solve the unrestricted problem due to the computation difficulties. Feature selection, therefore, is needed. The remaining subset of predictors, usually thought to be selected, are used to explain the different prevalences of disease between cases and controls.

2.1 The Lasso Penalty

The Least Absolute Shrinkage and Selection Operator (lasso) proposed by [Tibshirani \(1996\)](#) is a powerful technique for model selection and estimation in linear regression models. Adding an L_1 type penalty on the regression coefficients to the sum of squared residuals tends to produce sparse models. Consider the common Gaussian linear regression model

$$y_i = \mu + \sum_{j=1}^p x_{ij}\beta_j + \epsilon_i, i = 1, \dots, n$$

where $\epsilon = (\epsilon_1, \dots, \epsilon_n)^t \sim N(0, \sigma^2 I_n)$ is the random error vector. The lasso estimate is the solution to

$$\min_{\theta} \left\{ \sum_{i=1}^n (y_i - \mu - x_i^t \beta)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| \right\}, \quad (2.1)$$

where $\theta = (\mu, \beta_1, \dots, \beta_p)^t$ and $\lambda > 0$ is a tuning parameter.

As shown by [Tibshirani \(1996\)](#), lasso gives a sparse interpretable model and has

excellent prediction accuracy. In linear regression with lasso penalty, when positive tuning parameter λ increases, the penalty term $\lambda \sum_{j=1}^p |\beta_j|$ will be more and more dominant. When $\lambda = 0$, the lasso penalty has no effect on estimation and full least square estimate will be the solution. When λ becomes larger and goes to ∞ , even a small shift of certain β_j from 0 will add a great value to the objective function. Given the fact that we're minimizing the objective function, the penalty will force more β_j to be 0 to achieve the minimum. Tibshirani (1996) also discussed the sparsity via geometry. One can write lasso penalized linear regression as

$$\begin{aligned} \min_{\theta} \quad & \sum_{i=1}^n (y_i - \mu - \sum_{j=1}^p x_{ij} \beta_j)^2 \\ \text{s.t.} \quad & \sum_{j=1}^p |\beta_j| \leq t, \end{aligned} \tag{2.2}$$

where $t > 0$ is a positive parameter that can control how many predictors to be selected. For any given $t \geq 0$, there exists a $\lambda \geq 0$ such that (2.1) and (2.2) have the same solutions, and vice versa. $\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2$ equals the quadratic form $(\beta - \hat{\beta})^t X^t X (\beta - \hat{\beta})$ (plus a constant), which is an elliptical contour centering at the ordinary least square (OLS) estimate $\hat{\beta}$. In the case when $p = 2$, the feasible region $\sum_{j=1}^p |\beta_j| \leq t$ is a rotated square with vertices on coordinate axes. There are four vertices in this case, each of which corresponds either β_1 or β_2 equals zero. The lasso solution is the first place that the contour touches the rotated square. When t is large enough, the ellipse will be entirely contained within the square and then the lasso solution will be the same with OLS. However there is a chance that, as provided by (a) in Fig 1.1, the elliptical contour touches the square at a vertex, corresponding to a zero coefficient. This zero coefficient usually don't occur in ridge regression, where the penalty is $\sum_{j=1}^p \beta_j^2 \leq t$. In (b) of Fig 1.1, there's no corner to touch. Therefore in ridge regression when t is small enough,

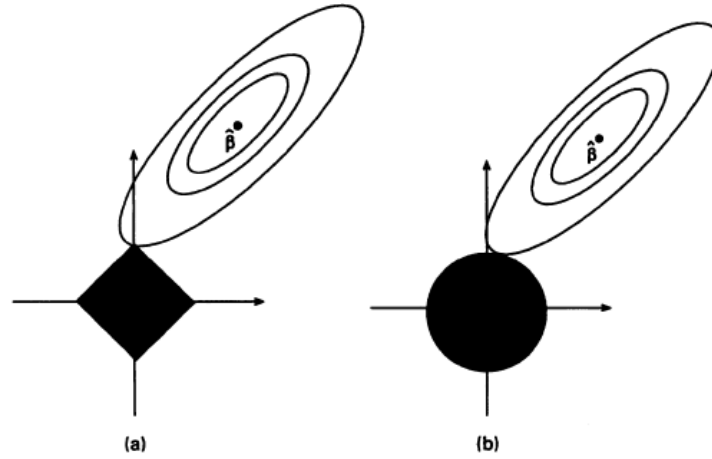


Figure 2.1: Estimate picture for (a) the lasso and (b) ridge regression (Tibshirani 1996)

coefficients could be compressed into small values but may never reach zeros.

2.2 The Lasso Estimate

One problem with lasso is that the objective function (2.2) is not differentiable and special optimization techniques are necessary. The lasso problem may be solved using quadratic programming or more general convex optimization methods, as well as by specific algorithms such as the least angle regression algorithm (Efron et al. 2004). Fu (1998) gave a shooting algorithm for lasso. The iteration starts with the OLS estimate and at step m the estimate is $\hat{\beta}^{(m)}$. At step $m + 1$ each component is updated as

$$\hat{\beta}_j^{(m+1)} = \begin{cases} [(\lambda - \partial g(\theta))/\partial \beta_j] / (2x_j^t x_j) & \text{if } \partial g(\theta)/\partial \beta_j > \lambda \\ [(-\lambda - \partial g(\theta))/\partial \beta_j] / (2x_j^t x_j) & \text{if } \partial g(\theta)/\partial \beta_j < -\lambda \\ 0 & \text{if } |\partial g(\theta)/\partial \beta_j| \leq \lambda \end{cases} .$$

It can be shown that $\hat{\beta}_j^{(m)}$ converges to the lasso estimate. However, Fu did not apply this algorithm to situations when the number of predictors far exceeded number of cases. One should note that this algorithm might not be applicable in situations where $p \gg n$. In

that case the starting point, the OLS estimate, may not be easy to calculate. [Daubechies et al. \(2004\)](#) re-discovered Fu's work by proposing an IST algorithm which could also be used to handle (2.2). IST only requires matrix-vector multiplications involving X and X^t ; convergence of IST algorithms was established by Daubechies and colleagues. Both Fu and Daubechies explicitly suggest coordinate descent for (2.2).

[Wu and Lange \(2008\)](#) applied a very fast and stable algorithm called cyclic coordinate descent for (2.2). The idea is to apply a coordinate-wise descent procedure for each value of the regularization parameter, varying the regularization parameter along a path. Each solution is used as a warm start for the next problem. This approach is attractive whenever the single-parameter problem is easy to solve. The update of the intercept parameter μ can be written as

$$\hat{\mu} = \frac{1}{n}(y_i - x_i^t \beta) = \mu - \frac{\partial}{\partial \mu} g(\theta).$$

For the parameter β_k , the update formulation depends on the direction:

$$\begin{aligned} \hat{\beta}_{k,+} &= \max \left\{ 0, \beta_k - \frac{\frac{\partial}{\partial \beta_k} g(\theta) + \lambda}{\sum_i x_{ik}^2} \right\} \\ \hat{\beta}_{k,-} &= \min \left\{ 0, \beta_k - \frac{\frac{\partial}{\partial \beta_k} g(\theta) - \lambda}{\sum_i x_{ik}^2} \right\}. \end{aligned}$$

At the same time, [Friedman et al. \(2007\)](#) also adopted the coordinate descent for lasso and fused lasso.

Later, [Wu et al. \(2009\)](#) discussed the cyclic coordinate ascent algorithm in logistic regression for variable selection in analysing case-control studies using SNPs. A score criterion was used to pre-select a working set of predictors to accelerate the search procedure. The tuning parameter λ can be selected for a fixed number of predictors using bracketing and the golden section search. In each update of β_k , Newton's method was

applied because the explicit maximum is not available. The lasso penalization proceeded in two stages. In the first stage, SNPs with strong main effects were identified; in the second stage, they looked for the interactions among the supported predictors.

2.3 Fused Lasso and Linear Structure of SNPs

As a generalization of lasso, fused lasso is designed for problems with features that can be ordered in some meaningful way ([Tibshirani et al. 2005](#)). The fused lasso problem is to solve

$$\begin{aligned} \min_{\theta} \sum_{i=1}^n (y_i - \mu - x_i^t \beta)^2 \\ \text{s.t. } \sum_{j=1}^p |\beta_j| \leq t_1 \text{ and } \sum_{j=2}^p |\beta_j - \beta_{j-1}| \leq t_2, \end{aligned} \quad (2.3)$$

where positive parameters t_1 and t_2 determine the strength of the two penalty term. We can always represent (2.3) in Lagrange form,

$$\min_{\theta} \left\{ \sum_{i=1}^n (y_i - \mu - x_i^t \beta)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^{p-1} |\beta_{j+1} - \beta_j| \right\}. \quad (2.4)$$

which is a strictly convex function. The first constraint is the usual lasso constraint encouraging sparse coefficients, as discussed above. Similar things will happen in fused lasso and the second encourages sparsity in their differences, that is, local constancy of the coefficient profile. When λ_2 increases, the fused lasso penalty $\lambda_2 \sum_{j=1}^{p-1} |\beta_{j+1} - \beta_j|$ will be more and more dominant. To minimize the objective function, we have to set more and more $\beta_{j+1} - \beta_j$ to 0 which means that the adjacent predictors are forced to be selected or dropped simultaneously.

The term ‘‘fusion’’ is borrowed from [Land and Friedman \(1997\)](#), which proposed the use of a penalty of the form $\sum_j |\beta_{j+1} - \beta_j|^\alpha \leq t_2$ for various values of α , especially

$\alpha = 0, 1, 2$. They did not consider the simultaneous use of both penalties $\sum_{j=1}^p |\beta_j|$ and $\sum_{j=1}^{p-1} |\beta_{j+1} - \beta_j|$ as in (2.3).

Coordinate-wise descent, however, does not work for the fused lasso (Friedman et al. 2007). For example, Proposition 2.7.1 of Bertsekas et al. (1999) shows that every limit point of successive coordinate-wise minimization of a continuously differentiable function is a stationary point for the overall minimization, provided that the minimum is uniquely obtained along each coordinate. Note that (2.4) is not continuously differentiable, implying that coordinate-wise descent can get stuck. Friedman et al. (2007) considered a variant of (2.4), called FLSA. For one-dimensional signals, they minimize

$$\frac{1}{2} \sum_{i=1}^p (y_i - \beta_i)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^{p-1} |\beta_{j+1} - \beta_j|. \quad (2.5)$$

A modified coordinate-wise algorithm was given for (2.5), which can be extended to the general fused lasso programs, but Friedman et al. (2007) don't guarantee the exact solution for general fused lasso problems. Liu et al. (2010) provided another possibility to solve fused lasso problems. Details are given in the next chapter.

2.4 Selection of Tuning Parameters

(Wu et al. 2009) adopted bracketing and golden section search. A pre-determined number of predictors could be selected with a given value of the tuning constant λ . Let $r(\lambda)$ be the number of predictors selected. If one reduces λ , the penalty will be relaxed, and more predictors can then enter the model. For every integer $s \leq p$, one could assume that there is an interval I_s on which $r(\lambda) = s$. A point in I_s can be found quickly by bracketing and bisection.

One can start with an estimate of λ in bracketing. If $r(\lambda) = s$, it's done. If $r(\lambda) < s$

and $a \in (0, 1)$ then there's a positive integer j such that $r(a^j \lambda) \geq s$. If $r(\lambda) > s$ and $b > 1$, then there's a positive integer k such that $r(b^k \lambda) \leq s$. One can set $a = .5$ and $b = 2$ and take the smallest integer j or k yielding the second bracketing point. Once one has a bracketing interval $[\lambda_l, \lambda_u]$, bisection is employed. This involves testing the midpoint $\lambda_m = (\lambda_l + \lambda_u)/2$. There are three possibilities: if $r(\lambda_m) = s$, it's done; if $r(\lambda_m) < s$, one replaces λ_u by λ_m ; otherwise, one replaces λ_l by λ_m . In either of the latter two cases, one bisects again and continues. The process is complete once we hit a point in I_s .

This method targeting a single tuning parameter works well to predetermine the number of selected predictors (Wu et al. 2009). In our problem we have two tuning parameters, hence this predetermined strategy is not that straightforward in our case. Instead we will select tuning variables based on validation method and BIC, as described in the following subsections.

Chapter 3

Method

SNPs do not function alone; it's very likely that adjacent ones function together and increase or decrease the likelihood of getting a particular disease. To incorporate such linear structures of SNPs, we include the map distance in the objective function. In addition to the lasso penalty, we also penalize on the fused lasso term weighted by the distance between adjacent SNPs. This double penalty forces the sparsity in solutions and the smoothness in adjacent SNPs with closer map distances as well. As a result, we expect to see those predictors altering risk because both high and low values can be selected.

In case-control samples, the response y_i is usually coded as 1 for cases and 0 for controls. The probability $p_i = P(y_i = 1)$ of getting a certain disease can be predicted by SNPs vector x_i . We model the logit function of p_i as

$$\text{logit}\left(\frac{p_i}{1-p_i}\right) = \mu + \sum_{j=1}^p x_{ij}\beta_j.$$

Usually we estimate the parameter vector $\theta = (\mu, \beta_1, \dots, \beta_p)^t$ by maximizing the loglikelihood function

$$L(\theta) = \sum_{i=1}^n [y_i \log p_i + (1 - y_i) \log(1 - p_i)].$$

To incorporate the linear structure of genetic networks, we add a lasso and weighted fused lasso penalty on $-L(\theta)$. Usually the predictor x_{ij} is set to 0, 1, 2, corresponding to the three SNPs genotypes aa, Aa and AA, respectively. One can always centralize x_{ij} by subtracting 1, so in the model we use x_{ij} taking values on $\{-1, 0, 1\}$.

With the double penalty, our objective function is (3.1), where ω_j is defined as

$1/(d_{j+1} - d_j)$, given that d_j is the local positions of SNP j . ω puts more weight between SNPs with closer distances. Biologically, SNPs from the same gene tend to cluster together and have closer distances between each other. SNPs from the same gene, therefore, tend to have similar function and should be present in the model at the same time (or to be discarded as the same time).

$$f(\theta) = -L(\theta) + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^{p-1} \omega_j |\beta_{j+1} - \beta_j| \quad (3.1)$$

The penalty part is not differentiable and not separable. Hence solving (3.1) is not straightforward. We will mimic the procedure used in the paper of Liu et al. (2010) to minimize (3.1). The procedure is combined from two algorithms: Nesterov’s method which is searching the estimate sequence and FLSA which can deal with a special case of fused lasso problem.

3.1 Nesterov’s Method

Nesterov’s method is an optimal first-order black-box method for smooth convex optimization (Nesterov 2003). It has been shown to improve the convergence properties of standard gradient-descent algorithms (Nesterov 1983). “Optimal” comes from the optimal convergence rate $O(1/k^2)$. In a classical gradient method, we take iteration

$$x^{(k+1)} = x^{(k)} - t_k g'(x^{(k)}) \quad (3.2)$$

or equivalently,

$$x^{(k+1)} = \operatorname{argmin}_x \left\{ g(x^{(k)}) + (x - x^{(k)})^t g'(x^{(k)}) + \frac{1}{2t_k} \|x - x^{(k)}\|^2 \right\} \quad (3.3)$$

with chosen $x^{(0)}$ and $\{t_k\}$ to minimize a convex differentiable function g . Here t_k is called the step size. Different step-size strategies include constant step, full relaxation and

Goldstein-Armijo rule (Nesterov 2003). The convergence rate of scheme (3.2) is $O(1/k)$ and Nesterov (1983) proved that there is room to improve $1/k$ rate of gradient method to $1/k^2$. Choosing $x^{(0)}$, set $s^{(0)} = x^{(0)}$ and $a_{-1} = a_0 = 1$, one can repeat the following Nesterov's improved steps (3.4) until convergence.

$$\begin{aligned} b_k &= (a_{k-2} - 1)/a_{k-1}, s^{(k)} = x^{(k)} + b_k(x^{(k)} - x^{(k-1)}) \\ x^{(k+1)} &= \operatorname{argmin}_x \left\{ g(s^{(k)}) + (x - s^{(k)})^t g'(s^{(k)}) + \frac{1}{2t_k} \|x - s^{(k)}\|^2 \right\} \\ a_{k+1} &= \left(1 + \sqrt{1 + 4a_k^2} \right) / 2 \end{aligned} \quad (3.4)$$

This scheme (3.4) has the optimal convergence rate $O(1/k^2)$. $\{x^{(k)}\}$ is the sequence of guessed solution. The novelty in (3.4) is that the sequence $\{s^{(k)}\}$ “remembers” the previous iterations through properly chosen a_k, b_k . And this $s^{(k)}$ makes the difference between standard gradient method and Nesterov's first-order method. Note that in standard gradient method (3.3) at k^{th} step, updating $x^{(k+1)}$ only depends on the current estimate $x^{(k)}$, while in the improved strategy by Nesterov it depends on $s^{(k)}$ which is the affine combination of current estimate $x^{(k)}$ and previous estimate $x^{(k-1)}$.

Nesterov's method has been further developed to deal with non-smooth function including lasso problem (Nesterov 2007). To our knowledge, it's still not able to deal with fused lasso problem.

3.2 Fused Lasso Signal Approximator

Friedman et al. (2007) modified the coordinate-wise descent procedure to minimize (3.5). Such modification is due to non-separable $|\beta_{i+1} - \beta_i|$.

$$g(\theta) = \frac{1}{2} \sum_{i=1}^p (y_i - \beta_j)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^{p-1} |\beta_{j+1} - \beta_j|. \quad (3.5)$$

Here we use notation g again. This modified procedure can be summarized into three nested cycles:

- *Descent cycle.* Coordinate-wise descent for each β_j . Set $\partial g(\beta)/\partial \beta_j = 0$. Update β_j as the solution. If cannot find the solution, examine $0, \beta_{j-1}, \beta_{j+1}$ and update β_j as the one giving the smallest value of $f(\beta)$.
- *Fusion cycle.* Enforce $|\beta_j - \beta_{j-1}| = 0$. That is, set $\beta_j = \beta_{j-1} = \gamma$ and reduce the problem to one with $p - 1$ parameters. At the end of descent and fusion cycles, identify adjacent non-zero and equal parameters, then collapse the data. Set $\beta_{j-1} = \beta_j = \gamma$ and $\partial g(\beta)/\partial \gamma = 0$. If the solution decreases $g(\beta)$, update β_{j-1} and β_j as γ .
- *Smoothing cycle.* Fix λ_1 , increase λ_2 from 0. δ is a small positive number.
 1. Start with $\lambda_2 = 0$.
 2. $\lambda_2 = \lambda_2 + \delta$. Run descent cycle and fusion cycle, till no further changes occur (terminate when change is less than certain threshold). After convergence, identify equal and non-zero neighbouring parameters and collapse the data.
 3. Repeat 2 until a target value of λ_2 is reached.

After updating the objective function, repeat descent and fusion cycles. The updating will be generally the same.

Collapsing data is to combine adjacent non-zero and equal parameters, assign weight to the observations averages and the contributions to the lasso penalty. After m fusions, the objective function has the form

$$\tilde{g}(\beta) = C_m + \frac{1}{2} \sum_{i=1}^{p-m} w_i (y_i - \beta_i)^2 + \lambda_1 \sum_{j=1}^{p-m} w_j |\beta_j| + \sum_{j=1}^{p-1-m} |\beta_{j+1} - \beta_j|.$$

Initially, $m = 0, w_i = 1, C_0 = 0, \tilde{g}(\beta) = g(\beta)$. C_m is irrelevant to β , we don't need to update it. If $(m + 1)^{\text{st}}$ fusion is between β_{i-1} and β_i , update the objective function as

- $\bar{y} = (w_{i-1}y_{i-1} + w_i y_i)/(w_{i-1} + w_i), \bar{w} = w_{i-1} + w_i$
- $y_{i-1} = \bar{y}, w_{i-1} = \bar{w}$
- Discard observation i . Update $i' = i' - 1$ if $i' > i$.

3.3 Nested Procedure of Nesterov's Method and FLSA

Friedman et al. (2007) clearly stated that FLSA cannot guarantee a desired solution for general fused lasso problem. Therefore to use FLSA, we should at least reform our objective function. On the other hand, Nesterov's scheme (3.4) provides an optimal first-order strategy in finding the sequence of solution while doesn't take care of the non-smooth parts $\sum_{j=1}^p |\beta_j|$ and $\sum_{j=1}^{p-1} \omega_j |\beta_{j+1} - \beta_j|$. Note when applying this strategy to minimize (3.1), the second step in (3.4) happens to be a fused lasso problem that could be solve by FLSA.

Specifically in each iteration, we use Nesterov's idea to update search point $s^{(k)}$ as the affine combination of current and previous estimates of coefficients. Then we jump to FLSA algorithm to update estimate based on $s^{(k)}$ which is a minimization problem $\underset{\theta}{\operatorname{argmin}} f_{R,s^{(k)}}(\theta)$, where

$$f_{R,\gamma}(\theta) = -[L(\gamma) + (\theta - \gamma)^t L'(\gamma)] + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^{p-1} |\beta_{j+1} - \beta_j| + \frac{R}{2} \|\theta - \gamma\|^2. \quad (3.6)$$

Constant R is chosen in each iteration according to the Goldstein-Armijo rule so that it should be appropriate for $s^{(k)}$. After updating the estimate, we then go back to Nesterov's scheme to update some constants and search next affine combination $s^{(k+1)}$. The whole

procedure is nested between Nesterov's method in dealing with smooth part and FLSA algorithm in non-smooth part. The whole procedure can be implemented in the following steps:

- Input: $\lambda_1, \lambda_2, \gamma_0, R_0$
- Initialize $\gamma_1 = \gamma_0, a_{-1} = 0, a_0 = 1, R = R_0$
- Loop and update
 1. $b_i = (a_{i-2} - 1)/a_{i-1}, s_i = \gamma_i + b_i(\gamma_i - \gamma_{i-1})$
 2. Find the smallest $R = R_{i-1}, 2R_{i-1}, \dots$ such that

$$f(\gamma_{i+1}) \leq f_{R, s_i}(\gamma_{i+1})$$

where $\gamma_{i+1} = \underset{\theta}{\operatorname{argmin}} f_{R, s_i}(\theta)$

3. Set $R_i = R, a_{i+1} = \left(1 + \sqrt{1 + 4a_i^2}\right) / 2$
4. End loop if

$$f(\gamma_k) - f(\gamma_{k+1}) \leq \frac{2 \max(2\tilde{R}, R_0) \|\gamma_0 - \gamma_{k+1}\|^2}{k^2}$$

where \tilde{R} is the Lipschitz continuous gradient of the negative loglikelihood function $-L(\cdot)$.

- γ_{k+1} is the optimal solution to (3.1).

3.4 Selection of Tuning Variables

3.4.1 Validation Method

Tuning parameters (λ_1, λ_2) can be selected based on validation method. We perform an exhaustive search through (a subset of) the space of (λ_1, λ_2) . The goal is to find the

optimal pair of (λ_1, λ_2) for the dataset. After selection on training data, we re-estimate the predictors without penalty. With estimation applied on tuning data, we record $L(\cdot)$, the loglikelihood function. The best pair (λ_1, λ_2) is determined following the greatest value of $L(\cdot)$. However, it may be possible that $L(\cdot)$ will continue to increase by adding more parameters into the model.

3.4.2 BIC Method and Its Extensions

One may also want to search the tuning parameters based on the Bayesian information criteria (BIC), which was introduced by [Schwarz \(1978\)](#) as a model selection tool. BIC is defined as (3.7), where n is the number of observations and k is the number of parameters. In part based on likelihood function, BIC also adjusts for the model size and prevents from overfitting. A smaller value represents a better fit. One drawback of BIC is that it tends to select a model with many spurious covariates.

$$\text{BIC} = -2L + k \times \ln(n) \tag{3.7}$$

There are some extensions of BIC. BICC, defined as (3.8), is one modification. Here p is the number of potential predictors.

$$\text{BICC} = -2L + k \times \ln(n) \times \max(1, \log(\log(p))) \tag{3.8}$$

Another option is the extended Bayesian information criteria (EBIC), defined as (3.9). First proposed by [Chen and Chen \(2008\)](#), EBIC can tightly control the false discovery rate in GWAS.

$$\text{EBIC} = -2L + k \times [\ln(n) + 2 \log(p)] \tag{3.9}$$

In simulation studies, each of these four measurements will give one pair of (λ_1, λ_2) . We will see how they perform in simulation studies.

Chapter 4

Analysis of Simulated Data

4.1 Input Setting of Simulated Data

We evaluate the performance of doubly penalized regression with linear structure with a focus on an underdetermined problem where p far exceeds n . We focus on the simulation model (4.1)

$$\log\left(\frac{p_i}{1-p_i}\right) = \mu + \sum_{j=1}^p x_{ij}\beta_j \quad (4.1)$$

Each predictor vector X_i , representing the DNA sequence variations, is generated from a realization of a multivariate normal vector Z_i whose margin is normal with mean 0, variance 0.1 and whose covariances are

$$\text{Cov}(Z_{ij}, Z_{ik}) = \begin{cases} 0.1 & j = k \\ 0.1 \times \rho^{|j-k|} & j, k \leq 5, j \neq k \\ 0.1 \times \rho^{|j-k|} & 6 \leq j, k \leq 10, j \neq k \\ 0.1 \times \rho^{|j-k|} & 11 \leq j, k \leq 20, j \neq k \\ 0 & \text{otherwise.} \end{cases}$$

which means that the predictors can be considered as four groups: predictors 1–5, 6–10, 11–20, and the rest. We set x_{ij} equal to -1 , 0 , or 1 when $Z_{ij} < -c$, $-c < Z_{ij} < c$, or $Z_{ij} > c$, respectively, where $c = 0.41$, the 2/3 quantile of the standard normal distribution. In every simulation, $\mu = 1$, $(\beta_1, \beta_2, \beta_7, \beta_9, \beta_{21}, \beta_{22}) = (1, 2, -1.5, -1, 1, -1)$, $\beta_j = 0$ for other j .

The value of $\omega_j = 1/(d_{j+1} - d_j)$ is used if the model incorporates the linear structure,

where d_j is the local position of SNP j . If all the entries of ω are equal, that means the model is without linear structure and locations of SNPs are equally spaced on the chromosome. In the model with linear structure to mimic the real situation, d is set in the following way,

$$d_j = \begin{cases} U_1 + d_{j-1} & j \leq 4 \text{ or } 6 \leq j \leq 9, \text{ or } 11 \leq j \leq 19 \\ U_2 + d_{j-1} & \text{otherwise} \end{cases} \quad (4.2)$$

where $U_1 \sim U(0.5, 1)$, $U_2 \sim U(5, 10)$. It implies that for first three groups, the SNPs within the same group are located closely. While the distances in last group and inter groups are relatively larger.

We carry out three penalties: and (1) lasso penalty, (2) double penalty without linear structure and (3) double penalty with linear structure. (1) only focuses on sparsity in SNPs; (2) treats SNPs equally located on chromosomes; (3), our proposed method, takes the distances between adjacent SNPs into consideration. The tuning parameters will be determined in a grid search based on the performance returned by greatest unpenalized loglikelihood in the validation, smallest BIC, BICC and EBIC. On the one hand, we are trying to determine which penalty is doing a better job selecting variables while on the other hand, we are comparing the four measurements in the grid search to decide which one is best for real data analysis.

4.2 Simulation Study Results

Results of 100 random samples are as demonstrated in Table 1 to Table 12 in the Appendix. Table 1 to Table 4 are generated from uncorrelated simulation data with $\rho = 0$, Table 5 to Table 8 from correlated data with $\rho = 0.5$, and Table 9 to Table 12 from $\rho = 0.9$. LS is short for “linear structure”. In each table, the first column (n, p) represents number

of observations and number of SNPs; the “Distance” column indicates the penalty setting; the third column is the average of determined tuning parameters; and column β_j shows the frequency of being selected in the 100 replications. For each distance setting, there are two rows of N and quantiles characterizing how “well” the corresponding distance setting is performing in selecting potential predictors. N_{true} refers to selected number of true predictors and $N_{\text{non-zero}}$ indicates the number of selected predictors. Monte Carlo sampling errors are in parenthesis. The last three columns record the training, tuning and testing error, separately. For example, in Table 1, when $(n, p) = (100, 50)$ with lasso penalty, the determined tuning parameters are $(0.039, 0)$. β_9 has the smallest frequency of 24 among 100 random samples. The average number of true positives is 2.6, which is not a very acceptable value. On average, about 4 predictors are selected.

When comparing results across the three penalties, we are looking for large true positives and smallest false positives because we would like to include as many true predictors into the model as possible and keep selection as clean as possible. We also prefer less testing error. In doing such comparison, one can do a simple test with data given in tables. Standard error $\sqrt{S_{\hat{N}}^2}$ is given in parenthesis. Here the Monte Carlo sampling error $S_{\hat{N}}^2 = \sum_{r=1}^R (\hat{N}_r - \hat{N})^2 / R$, given R random samples. When comparing two situations N_1 and N_2 , $\text{Var}(\hat{N}_1 - \hat{N}_2) \approx (S_{\hat{N}_1}^2 + S_{\hat{N}_2}^2) / R = \hat{V}_D$ is the estimated variance of the difference. We can draw the conclusion by looking at

$$z = \frac{\hat{N}_1 - \hat{N}_2}{\sqrt{\hat{V}_D}} \approx N(0, 1), \quad (4.3)$$

where large value of $|z|$ suggests significance. For example in the last block in Table 10 when comparing without linear structure and the proposed method, we have $|z| = |4.1 - 3.9| / \sqrt{(0.012^2 + 0.012^2)} \approx 11.9$. This large $|z|$ value indicates the significant difference

between the two methods, indicating a significant larger correctly number of selected true predictors.

When data is not correlated, EBIC is the most conservative method with the least predictors selected. Across all cases, the numbers of false positive are zero or close to zero but this doesn't mean that the results are acceptable since the total selected is always kept at a very low value. Under the validation method, BIC, there are one or two cases when (3) double penalty with linear structure is the best in terms of true positives. Specifically, when data is not correlated or when $\rho = 0.5$, the proposed method does not show advantage over the other two methods in our simulation studies. The proposed method, in a few cases, is doing as good as without linear structure. When data is correlated with $\rho = 0.9$, the proposed method is doing the best in all blocks of Table 10 and 12, in last two blocks of Table 11. On the other hand, we should notice that BIC tend to generate larger false positives. This may due to relatively small sample size. When sample size increases, false positives generated by BIC are almost the same with other three criteria.

Overall, the results of simulated data analysis do not give solid evidence that (3) is a better method in selecting SNPs. The differences between (1)–(3) are not very obvious. Higher correlation may increase the number of selected SNPs and the number of true positive as well. It's notable that lasso penalty seems to be robust against the four measurements of the validation, BIC, BICC and EBIC. The selection results of lasso do not change much in all cases. Among the four measurements, BIC may be a reasonable way to determine tuning parameters in the next step, the real data analysis.

Chapter 5

Real Data Analysis

Coeliac disease is an immune-mediated, chronic digestive disease that primarily affects the gastrointestinal tract and interferes with absorption of nutrients from food (Losowsky 2008). It can be triggered by gluten ingestion while gluten is a protein that can be found in wheat, rye, and barley. Symptoms include chronic diarrhoea, failure to thrive (in children), and fatigue (Rodrigo 2006). Among U.S. adults, the prevalence of coeliac disease is about 1% (Fasano et al. 2003; Rewers 2005). Several genetic factors combined with an environmental trigger are necessary for the disease to develop.

The purpose of real data analysis is to apply the proposed method on coeliac data and evaluate the feasibility and performance of the proposed method in real data. Though the association between HLA region and coeliac disease has been well established (van Heel and West 2006), van Heel et al. (2007) estimated that HLA region only contributes to 35% genetic variation. Wolters and Wijmenga (2008) also claimed that there should exist other non-HLA regions contributing to coeliac development. Both van Heel et al. (2007) and Wu et al. (2009) found that there are other strongly significant SNPs outside the HLA region. Besides the well known HLA-DQ2 and HLA-DQ8 located on chromosome 6p21, KIAA1109-TENR-IL2-IL21 block (van Heel et al. 2007), TNFAIP3 and REL (Trynka et al. 2009) were found to be new susceptibility factors for coeliac disease. Wu et al. (2009) identified some significant SNPs located on other chromosomes including chromosome 8. The coeliac disease data, therefore, is a good option to validate our proposed method. Due to the computation speed, we only look at chromosome 8.

5.1 Data Description

In the British coeliac data of [van Heel et al. \(2007\)](#), a number of 310637 SNPs were typed on $n = 2200$ subjects (938 males and 1262 females). The number of controls, defined as without coeliac disease, is 1422 while the number of cases, defined as with coeliac diseases, is 778. We only examined chromosome 8, which has a SNP number of 17904. Response is 1 if the subject is with coeliac disease and 0 otherwise. x_{ij} is set to $-1, 0, 1$, corresponding to the three SNPs genotypes aa, Aa and AA, respectively. All the SNPs are ordered; missing genotypes are imputed. The BP position as a measurement of local position will be used to calculate ω in the double penalty with linear structure. SNPs with closer BP position gain a greater weight in the double penalty.

We applied BIC to select tuning parameters because it showed a slight advantage over the lasso and double penalty without mapping distance across all the sample sizes and correlation settings in simulated data analysis. First we employed grid search to determine the tuning variables using BIC; next, with the selected λ_1 and λ_2 , 17904 SNPs as potential predictors were thrown into the model and the strongest predictors were selected. We identified those selected SNPs are the most “important” predictors. The selected SNPs, with the gender covariate, were then estimated again in logistic regression. P-values were recorded.

5.2 Results of Real Data Analysis

Table 13 summarizes the real data analysis results. The first column is the SNP marker name, except for the first row which shows information for gender; BP position indicates the base pair location on the chromosome. The column p-value is copied from

the logistic regression analysis after selection; the columns of gene and function record the gene name that the corresponding SNP belongs to and its related functions, if applicable.

The proposed method reduced the number of potential predictors on chromosome 8 from 17904 to 28. Among the 28 selected SNPs, 13 are intergenic SNPs between other genes. Occasionally, intergenic regions act to control genes nearby, but most of them have no currently known function. There are 13 SNPs belonging to 12 genes. We were unable to find any information for SNP rs10503561 and rs7386962.

Specifically, both SNPs rs6995469 and rs6991080 are from gene CSMD1, which is a potential suppressor of squamous cell carcinomas ([Lau and Scholnick 2003](#)). CSMD1 has been found to be a potential risk factor of schizophrenia ([Håvik et al. 2011](#)) and chemical dependency ([Rose et al. 2010](#)). SNP rs3943520 is included in gene SLC39A14, which is zinc transporter ([Taylor et al. 2005](#)). Zinc is an important cofactor for many enzymes. SNP rs1485750 locates in gene EBF2. As a protein-coding gene, EBF2 is a transcription factor that, in osteoblasts, activates the decoy receptor for RANKL, TNFRSF11B, which in turn regulates osteoclast differentiation and acts in synergy with the Wnt-responsive LEF1/CTNNB1 pathway ([Wang et al. 2002](#)). SNP rs2347501 is from gene NRG1 encoding one of four proteins in the neuregulin family that act on the EGFR family of receptors. It is important for the normal development of the nervous system and the heart ([Peles et al. 1992](#); [Plowman et al. 1993](#)). NRG1 may also be an important factor in developing schizophrenia ([Zhao et al. 2004](#)) and breast tumors ([Huang et al. 2004](#)). Gene C80rf72, which rs10504244 belongs to, may play certain role in tumor progression ([Hauge et al. 2007](#)).

SNPs rs1445401, rs648119, rs10096287, rs4871072, rs10505604 and rs2896714 belong to gene EYA1, NCALD, RIMS2, SNTB1, TG and ST3GAL1 separately. These genes also

have their biological functions. EYA1 encodes a member of the eyes absent (EYA) family of proteins. The encoded protein may play a role in the developing kidney, branchial arches, eye, and ear. Mutations of this gene have been associated with branchiootorenal dysplasia syndrome, branchiootic syndrome, and sporadic cases of congenital cataracts and ocular anterior segment anomalies (Cook et al. 2009). NCALD may be involved in the calcium-dependent regulation of rhodopsin phosphorylation (Alexanian et al. 2001). RIMS2 is a rab effector involved in exocytosis and may act as scaffold protein (Wang and Südhof 2003). SNTB1 encodes a large, rod-like cytoskeletal protein found at the inner surface of muscle fibers, Dystrophin. The protein encoded by this gene is a peripheral membrane protein that has been found to be associated with dystrophin and dystrophin-related proteins. This gene is a member of the syntrophin gene family, which contains at least two other structurally-related genes (Ahn and Kunkel 1995). It's reported that TG is working as a thyroid hormone precursor, storage of iodine, and storage of inactive thyroid hormones (Boat et al. 1989, p. 1854-1861). The protein encoded by ST3GAL1 is normally found in the Golgi but can be proteolytically processed to a soluble form. Correct glycosylation of the encoded protein may be critical to its sialyltransferase activity (Kitagawa and Paulson 1994). The association between ST3GAL1 and bipolar disorder have been reported (Zandi et al. 2008; Zhang et al. 2010). We were not able to collect any biological function of gene SAMD12 or confirmed function on absorption and digestion of selected genes.

Looking into our results, our findings and the current literature do have some overlapping results. In GWAS of Wu et al. (2009) using the same data set, two intergenic SNPs rs736191 and rs1499447, and SNP rs10505604 in gene TG were reported to be strongly associated with coeliac disease. On the other hand, Trynka et al. (2009) found

there is no association between coeliac disease and the gene CSMD1, C8orf72, EYA1 and TG. To our knowledge, there is no other literature on association between the other genes and digestion (or absorption).

Chapter 6

Conclusion

Penalized logistic regression is applied to model selection to handle “large p , small n ” gene mapping problems. The proposed method explicitly incorporates a measure of linear structure of adjacent SNPs to encourage smoothness of the effects of those with close local positions.

The simulation analysis shows that this double penalty with linear structure is superior to that without linear structure and the lasso penalty in SOME cases in terms of both the numbers of true positives and false positives. By forcing the fusion of adjacent SNPs weighted by difference of their local positions, the proposed method can identify the strongest associated SNPs among the candidate pool. The other two methods, either fail to consider the smoothness of the adjacent SNPs or don’t take the linear structure into account. Thus, the variants raising the risk by small values may not be detected. By imposing the linear structure, two adjacent SNPs with closer local position are fused with a greater weight and are simultaneously more likely to be selected or to be dropped.

The proposed method, however, is not always an ideal option as there is certain chance that relatively weak associated predictors may not be chosen. Another thorny issue in handling large number of SNPs simultaneously is computation. We borrowed the Nesterov method to solve the objective function, but the computation speed is under expectations especially when the sample size is large. Simulation studies also show that the discrepancies between our proposed method and the other two methods are not very obvious. Originally we proposed to apply coordinate descent to this problem (with some

modifications by replacing the fused lasso term with a total variation norm). Coordinate descent is applicable when the problem is separable. Our modification did not avoid the non-separability thus coordinate descent did not work in our problem.

In the coeliac data analysis, we focused on chromosome 8. The whole procedure can therefore be generalized to the whole genome. With the 28 selected SNPs, we identified 12 genes, including the gene TG. In GWAS of [Wu et al. \(2009\)](#) using the same data set, two intergenic SNPs rs736191 and rs1499447, and SNP rs10505604 in gene TG were also reported as the significant markers on chromosome 8. On the other hand, [Trynka et al. \(2009\)](#) found that TG was not associated to coeliac disease. Our proposed method detected statistically significant predictors related to coeliac disease but due to computational speed, our analysis was restricted within chromosome 8. A similar procedure using our proposed method on the whole genome may be able to identify more biologically significant genes.

Table 1: Simulation results for $\rho = 0$ based on validation method

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.039, 0.000)	30	84	61	24	26	39	2.6(0.012) (1.0, 3.0, 4.0)	4.1(0.023) (2.0, 4.0, 7.0)	21.61%	29.98%	29.47%
	w/o LS ^a	(0.030, 0.014)	71	92	66	30	27	29	3.1(0.013) (2.0, 3.0, 5.0)	6.1(0.040) (2.0, 5.0, 12.0)	20.90%	29.92%	30.01%
	w/ LS	(0.029, 0.035)	60	82	54	19	37	35	2.9(0.014) (1.0, 3.0, 5.0)	7.7(0.062) (2.0, 6.5, 16.0)	21.04%	30.56%	30.64%
(100, 500)	Lasso	(0.045, 0.000)	15	65	44	11	13	20	1.7(0.010) (1.0, 2.0, 3.0)	8.2(0.032) (5.0, 8.0, 11.0)	12.93%	36.54%	36.65%
	w/o LS	(0.034, 0.023)	66	81	33	13	3	9	2.0(0.011) (1.0, 2.0, 4.0)	6.2(0.067) (2.0, 5.0, 12.5)	19.30%	33.10%	33.47%
	w/ LS	(0.046, 0.025)	19	54	34	8	11	22	1.5(0.010) (0.0, 1.0, 3.0)	8.6(0.032) (5.0, 8.0, 12.0)	13.26%	37.35%	37.27%
(200, 500)	Lasso	(0.043, 0.000)	26	91	57	17	25	21	2.4(0.011) (1.0, 2.0, 4.0)	3.5(0.020) (1.5, 3.0, 6.0)	24.18%	28.77%	28.65%
	w/o LS	(0.025, 0.013)	80	99	72	42	30	26	3.5(0.012) (2.0, 4.0, 5.0)	8.9(0.059) (3.0, 8.0, 17.5)	20.67%	29.68%	29.74%
	w/ LS	(0.042, 0.010)	32	85	54	17	25	21	2.3(0.011) (1.0, 2.0, 4.0)	3.8(0.023) (1.0, 3.0, 7.0)	24.16%	29.38%	29.22%
(200, 1000)	Lasso	(0.044, 0.000)	24	88	54	13	17	15	2.1(0.010) (1.0, 2.0, 3.0)	3.6(0.020) (2.0, 3.0, 6.0)	23.52%	29.91%	29.72%
	w/o LS	(0.027, 0.014)	79	97	66	40	22	21	3.2(0.011) (2.0, 3.0, 5.0)	9.4(0.064) (3.0, 8.0, 19.0)	20.11%	30.91%	31.01%
	w/ LS	(0.044, 0.009)	31	79	50	11	19	16	2.1(0.011) (1.0, 2.0, 3.0)	3.7(0.021) (1.5, 3.0, 7.0)	23.73%	30.65%	30.62%
(500, 1000)	Lasso	(0.030, 0.000)	52	100	100	48	44	60	4.0(0.009) (3.0, 4.0, 5.0)	4.7(0.023) (3.0, 4.0, 6.0)	24.51%	26.18%	26.18%
	w/o LS	(0.019, 0.007)	92	100	100	80	72	64	5.1(0.008) (4.0, 5.0, 6.0)	14.2(0.088) (5.0, 11.0, 28.0)	21.34%	27.87%	27.18%
	w/ LS	(0.030, 0.016)	64	92	92	40	36	56	3.8(0.013) (3.0, 4.0, 6.0)	5.0(0.025) (3.0, 5.0, 8.0)	24.70%	26.77%	26.33%

^a“LS” refers to linear structure. The rest are the same.

Table 2: Simulation results for $\rho = 0$ based on BIC

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.037, 0.000)	31	77	57	28	29	41	2.6(0.012) (1.0, 2.0, 4.0)	4.9(0.038) (2.0, 4.0, 8.5)	19.56%	NA	30.97%
	w/o LS	(0.035, 0.002)	35	79	58	30	33	42	2.8(0.012) (1.0, 3.0, 4.0)	5.6(0.039) (2.0, 5.0, 9.0)	19.43%	NA	31.36%
	w/ LS	(0.036, 0.002)	35	80	59	27	35	41	2.8(0.013) (1.0, 3.0, 4.0)	5.6(0.051) (2.0, 4.0, 9.0)	19.41%	NA	30.85%
(100, 500)	Lasso	(0.038, 0.000)	23	74	47	17	16	31	2.1(0.011) (1.0, 2.0, 4.0)	14.8(0.048) (7.5, 16.0, 20.0)	3.80%	NA	38.03%
	w/o LS	(0.032, 0.004)	42	76	53	16	21	32	2.4(0.012) (1.0, 2.0, 4.0)	19.9(0.065) (10.0, 19.5, 28.0)	2.32%	NA	38.13%
	w/ LS	(0.038, 0.022)	30	66	42	14	17	28	2.0(0.011) (0.0, 2.0, 3.0)	16.6(0.059) (10.0, 16.0, 23.5)	3.68%	NA	38.84%
(200, 500)	Lasso	(0.028, 0.000)	54	98	88	51	55	56	4.0(0.012) (2.0, 4.0, 6.0)	19.0(0.116) (6.5, 14.5, 37.5)	11.23%	NA	32.83%
	w/o LS	(0.025, 0.002)	62	99	90	55	58	59	4.2(0.011) (3.0, 4.0, 6.0)	23.8(0.127) (10.0, 21.0, 44.0)	9.64%	NA	33.42%
	w/ LS	(0.027, 0.008)	57	97	86	48	57	61	4.1(0.012) (3.0, 4.0, 6.0)	23.1(0.126) (8.0, 21.5, 39.5)	9.49%	NA	33.80%
(200, 1000)	Lasso	(0.030, 0.000)	43	95	82	48	43	46	3.6(0.011) (2.0, 4.0, 5.0)	21.1(0.104) (8.0, 18.5, 36.0)	8.78%	NA	34.83%
	w/o LS	(0.025, 0.003)	54	100	89	50	50	56	4.0(0.010) (3.0, 4.0, 5.0)	33.6(0.147) (14.5, 32.0, 54.5)	4.43%	NA	35.28%
	w/ LS	(0.028, 0.016)	51	89	72	35	47	53	3.5(0.012) (2.0, 3.0, 5.0)	28.5(0.116) (11.5, 31.0, 40.0)	5.87%	NA	36.64%
(500, 1000)	Lasso	(0.022, 0.000)	72	100	100	84	84	72	5.1(0.008) (4.0, 5.0, 6.0)	15.0(0.067) (5.0, 15.0, 23.0)	19.60%	NA	28.42%
	w/o LS	(0.020, 0.001)	76	100	100	84	92	80	5.3(0.007) (5.0, 5.0, 6.0)	18.2(0.073) (12.0, 16.0, 33.0)	18.72%	NA	29.13%
	w/ LS	(0.022, 0.001)	72	100	100	84	84	72	5.1(0.008) (4.0, 5.0, 6.0)	14.0(0.047) (5.0, 15.0, 19.0)	19.96%	NA	28.16%

Table 3: Simulation results for $\rho = 0$ based on BICC

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.046, 0.000)	21	73	49	17	19	30	1.0(0.000)	1.0(0.000)	21.91%	NA	28.57%
	w/o LS	(0.046, 0.001)	25	73	51	18	24	32	(1.0, 1.0, 1.0)	(1.0, 1.0, 1.0)	21.53%	NA	28.57%
	w/ LS	(0.046, 0.001)	26	74	52	17	23	34	(1.0, 1.0, 1.0)	(1.0, 1.0, 1.0)	21.38%	NA	28.57%
(100, 500)	Lasso	(0.046, 0.000)	15	67	41	11	11	20	1.0(0.000)	8.0(0.000)	12.21%	NA	35.38%
	w/o LS	(0.041, 0.006)	25	65	39	9	11	16	(1.0, 1.0, 1.0)	(8.0, 8.0, 8.0)	15.12%	NA	35.50%
	w/ LS	(0.046, 0.001)	13	61	38	10	11	20	(1.0, 1.0, 1.0)	(9.0, 9.0, 9.0)	12.36%	NA	35.38%
(200, 500)	Lasso	(0.031, 0.000)	24	88	56	17	20	16	4.0(0.000)	5.0(0.000)	24.00%	NA	27.77%
	w/o LS	(0.031, 0.001)	26	91	62	21	26	22	(4.0, 4.0, 4.0)	(5.0, 5.0, 5.0)	22.71%	NA	27.77%
	w/ LS	(0.031, 0.001)	25	88	58	20	23	20	(4.0, 4.0, 4.0)	(5.0, 5.0, 5.0)	23.63%	NA	27.77%
(200, 1000)	Lasso	(0.041, 0.000)	21	86	50	11	17	14	3.0(0.000)	3.0(0.000)	23.44%	NA	26.16%
	w/o LS	(0.041, 0.001)	25	86	56	16	23	19	(3.0, 3.0, 3.0)	(3.0, 3.0, 3.0)	21.77%	NA	26.16%
	w/ LS	(0.041, 0.001)	22	83	54	12	17	16	(3.0, 3.0, 3.0)	(3.0, 3.0, 3.0)	23.27%	NA	26.16%
(500, 1000)	Lasso	(0.029, 0.000)	48	100	100	52	52	64	4.2(0.009)	4.7(0.023)	24.11%	NA	26.58%
	w/o LS	(0.028, 0.001)	52	100	100	64	48	72	(3.0, 4.0, 5.0)	(3.0, 4.0, 6.0)	23.42%	NA	26.63%
	w/ LS	(0.029, 0.003)	48	100	100	56	44	60	4.1(0.008)	5.0(0.025)	24.18%	NA	26.35%

Table 4: Simulation results for $\rho = 0$ based on EBIC

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.046, 0.000)	18	68	49	16	14	26	1.0(0.000)	1.0(0.000)	22.62%	NA	28.57%
	w/o LS	(0.046, 0.001)	25	63	36	9	8	14	(1.0, 1.0, 1.0)	(1.0, 1.0, 1.0)	24.31%	NA	28.57%
	w/ LS	(0.046, 0.001)	41	81	41	15	19	33	(1.0, 1.0, 1.0)	(1.0, 1.0, 1.0)	21.84%	NA	28.57%
(100, 500)	Lasso	(0.046, 0.000)	14	64	41	11	11	19	1.0(0.000)	8.0(0.000)	13.18%	NA	35.38%
	w/o LS	(0.046, 0.021)	28	49	25	2	2	5	(1.0, 1.0, 1.0)	(8.0, 8.0, 8.0)	23.47%	NA	28.57%
	w/ LS	(0.046, 0.001)	11	59	38	10	11	20	(1.0, 1.0, 1.0)	(1.0, 1.0, 1.0)	13.21%	NA	35.38%
(200, 500)	Lasso	(0.041, 0.000)	18	85	46	9	14	11	3.0(0.000)	3.0(0.000)	25.00%	NA	26.16%
	w/o LS	(0.041, 0.006)	24	83	40	10	11	9	(3.0, 3.0, 3.0)	(3.0, 3.0, 3.0)	25.29%	NA	26.16%
	w/ LS	(0.041, 0.001)	14	80	40	8	15	9	(3.0, 3.0, 3.0)	(3.0, 3.0, 3.0)	25.25%	NA	26.16%
(200, 1000)	Lasso	(0.041, 0.000)	18	84	46	9	14	11	3.0(0.000)	3.0(0.000)	24.20%	NA	26.16%
	w/o LS	(0.041, 0.006)	25	77	39	9	10	8	(3.0, 3.0, 3.0)	(3.0, 3.0, 3.0)	24.98%	NA	26.16%
	w/ LS	(0.046, 0.011)	18	78	40	8	13	10	(3.0, 3.0, 3.0)	(3.0, 3.0, 3.0)	24.64%	NA	26.16%
(500, 1000)	Lasso	(0.037, 0.000)	24	100	76	16	16	24	2.6(0.010)	2.6(0.010)	25.66%	NA	26.96%
	w/o LS	(0.035, 0.001)	24	100	76	24	28	36	(1.0, 3.0, 4.0)	(1.0, 3.0, 4.0)	25.35%	NA	26.84%
	w/ LS	(0.035, 0.001)	24	100	76	24	24	28	2.9(0.010)	3.1(0.012)	25.46%	NA	27.02%

Table 5: Simulation results for $\rho = 0.5$ based on validation method

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.040, 0.000)	42	81	66	21	30	27	3.0(0.010)	4.2(0.015)	21.20%	29.07%	27.93%
	w/o LS	(0.033, 0.021)	69	84	66	39	21	21	3.3(0.013)	(2.0, 3.0, 4.5) 5.8(0.038)	21.20%	28.17%	28.00%
	w/ LS	(0.037, 0.039)	60	81	54	30	24	30	3.1(0.013)	(2.0, 3.0, 5.0) 6.7(0.053)	20.50%	29.93%	29.37%
(100, 500)	Lasso	(0.045, 0.000)	33	69	45	18	12	21	2.2(0.010)	8.5(0.023)	13.77%	36.30%	35.67%
	w/o LS	(0.036, 0.024)	69	81	39	30	9	12	(1.0, 2.0, 3.5) 2.7(0.012)	(5.5, 8.0, 11.5) 7.2(0.058)	18.67%	32.03%	31.40%
	w/ LS	(0.045, 0.017)	36	66	48	21	9	18	(1.0, 3.0, 4.0) 2.2(0.010)	(3.0, 5.0, 15.5) 9.5(0.033)	13.13%	36.83%	35.13%
(200, 500)	Lasso	(0.044, 0.000)	36	100	72	32	32	16	2.9(0.010)	3.7(0.016)	23.18%	27.04%	26.66%
	w/o LS	(0.029, 0.015)	76	100	76	52	36	20	(2.0, 3.0, 4.0) 3.6(0.011)	(2.0, 3.0, 6.0) 7.6(0.051)	20.88%	28.20%	27.78%
	w/ LS	(0.044, 0.015)	48	100	72	28	24	16	(2.0, 4.0, 5.0) 2.9(0.009)	(3.0, 6.0, 19.0) 4.1(0.016)	23.30%	27.32%	27.18%
(200, 1000)	Lasso	(0.045, 0.000)	32	96	72	28	20	16	2.6(0.009)	3.8(0.017)	22.60%	28.28%	27.96%
	w/o LS	(0.028, 0.019)	76	100	72	52	20	20	(1.0, 3.0, 4.0) 3.4(0.011)	(2.0, 3.0, 6.0) 7.6(0.049)	20.60%	28.72%	28.86%
	w/ LS	(0.044, 0.007)	40	100	72	28	24	16	(2.0, 3.0, 5.0) 2.8(0.010)	(3.0, 6.0, 12.0) 4.4(0.023)	22.00%	28.66%	28.46%
(500, 1000)	Lasso	(0.030, 0.000)	32	40	32	40	16	24	4.6(0.014)	5.4(0.019)	24.28%	24.08%	24.92%
	w/o LS	(0.021, 0.005)	40	40	40	40	24	24	(3.0, 5.0, 6.0) 5.2(0.010)	(3.0, 5.0, 8.0) 14.0(0.086)	21.28%	26.56%	26.84%
	w/ LS	(0.033, 0.009)	32	40	32	40	16	16	(4.0, 6.0, 6.0) 4.4(0.014)	(7.0, 10.0, 30.0) 5.8(0.027)	24.20%	24.40%	25.16%

Table 6: Simulation results for $\rho = 0.5$ based on BIC

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.036, 0.000)	42	78	63	30	24	36	3.0(0.014)	6.3(0.057)	18.27%	NA	30.50%
	w/o LS	(0.034, 0.002)	42	81	66	30	33	39	3.2(0.014)	7.0(0.060)	17.57%	NA	29.93%
	w/ LS	(0.034, 0.004)	48	78	69	36	33	36	3.3(0.014)	7.6(0.066)	17.10%	NA	29.87%
(100, 500)	Lasso	(0.037, 0.000)	39	75	54	21	18	24	2.6(0.009)	16.2(0.047)	2.67%	NA	39.93%
	w/o LS	(0.033, 0.003)	36	75	57	27	18	30	2.7(0.009)	20.0(0.083)	2.83%	NA	38.77%
	w/ LS	(0.039, 0.016)	45	75	48	18	21	27	2.6(0.010)	16.5(0.069)	4.80%	NA	38.90%
(200, 500)	Lasso	(0.030, 0.000)	64	100	84	56	48	40	3.9(0.012)	16.4(0.105)	13.04%	NA	31.40%
	w/o LS	(0.026, 0.001)	72	100	88	60	60	48	4.3(0.010)	22.2(0.107)	10.06%	NA	32.60%
	w/ LS	(0.026, 0.017)	88	100	80	60	48	48	4.2(0.010)	24.2(0.115)	9.18%	NA	33.66%
(200, 1000)	Lasso	(0.032, 0.000)	56	100	88	56	48	24	3.7(0.010)	18.1(0.094)	10.72%	NA	31.98%
	w/o LS	(0.026, 0.002)	60	100	88	60	52	40	4.0(0.011)	31.0(0.164)	6.42%	NA	33.06%
	w/ LS	(0.029, 0.019)	56	96	84	56	52	24	3.7(0.010)	25.0(0.111)	7.94%	NA	34.14%
(500, 1000)	Lasso	(0.023, 0.000)	40	40	40	40	24	32	5.4(0.008)	13.8(0.053)	19.96%	NA	27.16%
	w/o LS	(0.020, 0.001)	40	40	40	40	24	32	5.4(0.008)	19.8(0.097)	18.24%	NA	28.08%
	w/ LS	(0.022, 0.016)	40	40	32	32	24	32	5.0(0.009)	15.0(0.039)	20.32%	NA	27.52%

Table 7: Simulation results for $\rho = 0.5$ based on BICC

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.042, 0.000)	39	75	57	18	18	27	2.6(0.011) (1.0, 3.0, 4.0)	4.2(0.015) (2.0, 4.0, 6.0)	21.60%	NA	28.90%
	w/o LS	(0.043, 0.003)	39	75	60	18	18	27	2.6(0.011) (1.0, 3.0, 4.0)	5.8(0.038) (2.5, 5.0, 10.5)	21.50%	NA	29.30%
	w/ LS	(0.042, 0.005)	39	72	57	18	21	27	2.6(0.011) (1.0, 3.0, 4.0)	6.7(0.053) (2.0, 5.0, 14.0)	21.97%	NA	29.60%
(100, 500)	Lasso	(0.045, 0.000)	33	66	45	18	12	21	2.2(0.010) (1.0, 2.0, 3.5)	8.5(0.023) (5.5, 8.0, 11.5)	13.03%	NA	36.00%
	w/o LS	(0.045, 0.008)	42	66	42	12	3	18	2.0(0.011) (1.0, 2.0, 3.0)	7.2(0.058) (3.0, 5.0, 15.5)	17.37%	NA	34.43%
	w/ LS	(0.045, 0.006)	36	66	45	18	12	21	2.2(0.011) (1.0, 2.0, 4.0)	9.5(0.033) (5.5, 9.0, 14.0)	12.77%	NA	36.50%
(200, 500)	Lasso	(0.045, 0.000)	28	96	64	32	20	16	2.6(0.009) (1.0, 3.0, 4.0)	3.7(0.016) (2.0, 3.0, 6.0)	23.66%	NA	26.94%
	w/o LS	(0.043, 0.001)	28	100	72	32	28	16	2.8(0.009) (2.0, 3.0, 4.0)	7.6(0.051) (3.0, 6.0, 19.0)	22.52%	NA	27.96%
	w/ LS	(0.043, 0.004)	32	100	68	32	24	16	2.7(0.008) (2.0, 3.0, 4.0)	4.1(0.016) (2.0, 4.0, 6.0)	22.50%	NA	28.12%
(200, 1000)	Lasso	(0.044, 0.000)	28	100	60	36	24	16	2.6(0.009) (1.0, 3.0, 4.0)	3.8(0.017) (2.0, 3.0, 6.0)	21.88%	NA	28.38%
	w/o LS	(0.044, 0.001)	32	100	60	36	24	16	2.7(0.009) (1.0, 3.0, 4.0)	7.6(0.049) (3.0, 6.0, 12.0)	21.92%	NA	28.30%
	w/ LS	(0.044, 0.003)	32	96	56	36	24	16	2.6(0.009) (1.0, 3.0, 4.0)	4.4(0.023) (2.0, 4.0, 8.0)	21.72%	NA	28.98%
(500, 1000)	Lasso	(0.032, 0.000)	32	40	32	40	8	16	4.2(0.010) (3.0, 5.0, 5.0)	5.4(0.019) (3.0, 5.0, 8.0)	24.56%	NA	25.64%
	w/o LS	(0.030, 0.001)	32	40	32	40	16	16	4.4(0.012) (3.0, 5.0, 6.0)	14.0(0.086) (7.0, 10.0, 30.0)	24.16%	NA	25.52%
	w/ LS	(0.031, 0.014)	32	40	32	40	16	16	4.4(0.012) (3.0, 5.0, 6.0)	5.8(0.027) (3.0, 4.0, 10.0)	24.60%	NA	25.20%

Table 8: Simulation results for $\rho = 0.5$ based on EBIC

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.043, 0.000)	39	72	57	18	15	24	2.5(0.011) (1.0, 3.0, 4.0)	3.3(0.016) (1.0, 3.0, 5.5)	21.87%	NA	27.90%
	w/o LS	(0.044, 0.022)	48	63	33	12	3	9	1.9(0.007) (1.0, 2.0, 3.0)	2.4(0.011) (1.0, 2.0, 4.0)	25.03%	NA	27.82%
	w/ LS	(0.042, 0.037)	51	72	33	24	12	15	2.3(0.010) (1.0, 2.0, 3.5)	4.3(0.029) (1.0, 4.0, 8.5)	23.10%	NA	28.31%
(100, 500)	Lasso	(0.046, 0.000)	33	66	45	18	12	21	2.2(0.010) (1.0, 2.0, 3.5)	8.3(0.024) (5.5, 8.0, 11.5)	13.73%	NA	35.67%
	w/o LS	(0.045, 0.035)	48	54	21	12	0	3	1.5(0.007) (1.0, 2.0, 2.0)	2.2(0.010) (1.0, 2.0, 3.5)	26.07%	NA	29.60%
	w/ LS	(0.046, 0.008)	39	63	45	18	12	21	2.2(0.010) (1.0, 2.0, 3.5)	8.5(0.023) (6.0, 8.0, 12.0)	13.60%	NA	35.80%
(200, 500)	Lasso	(0.046, 0.000)	28	96	60	28	16	16	2.4(0.009) (1.0, 3.0, 3.0)	3.0(0.013) (1.0, 3.0, 4.0)	24.00%	NA	26.70%
	w/o LS	(0.045, 0.012)	40	96	40	12	12	8	2.1(0.006) (1.0, 2.0, 3.0)	2.4(0.009) (1.0, 2.0, 4.0)	25.64%	NA	27.04%
	w/ LS	(0.046, 0.010)	28	96	48	20	20	16	2.3(0.007) (1.0, 2.0, 3.0)	2.9(0.011) (1.0, 3.0, 4.0)	24.48%	NA	27.40%
(200, 1000)	Lasso	(0.046, 0.000)	28	96	60	28	16	16	2.4(0.009) (1.0, 3.0, 3.0)	3.3(0.012) (2.0, 3.0, 5.0)	23.12%	NA	27.98%
	w/o LS	(0.046, 0.013)	40	92	36	12	8	8	2.0(0.006) (1.0, 2.0, 3.0)	2.4(0.011) (1.0, 2.0, 4.0)	25.18%	NA	27.72%
	w/ LS	(0.046, 0.014)	32	80	48	24	16	16	2.2(0.008) (1.0, 2.0, 3.0)	3.2(0.013) (1.0, 3.0, 5.0)	23.70%	NA	29.30%
(500, 1000)	Lasso	(0.039, 0.000)	24	40	24	16	0	0	2.6(0.005) (2.0, 3.0, 3.0)	2.6(0.005) (2.0, 3.0, 3.0)	25.96%	NA	26.80%
	w/o LS	(0.037, 0.001)	24	40	32	24	0	8	3.2(0.010) (2.0, 3.0, 5.0)	3.4(0.010) (2.0, 3.0, 5.0)	25.56%	NA	26.36%
	w/ LS	(0.037, 0.014)	24	40	32	24	0	8	3.2(0.010) (2.0, 3.0, 5.0)	3.6(0.010) (2.0, 4.0, 5.0)	25.56%	NA	26.36%

Table 9: Simulation results for $\rho = 0.9$ based on validation method

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.038, 0.000)	45	80	65	35	35	40	3.0(0.013) (1.5, 3.0, 5.0)	5.2(0.019) (3.0, 5.0, 7.5)	20.30%	29.25%	26.65%
	w/o LS	(0.032, 0.024)	70	85	75	65	30	30	3.5(0.014) (2.0, 3.5, 5.5)	8.1(0.051) (4.0, 6.5, 16.0)	20.60%	31.30%	28.50%
	w/ LS	(0.039, 0.068)	65	75	60	50	15	25	2.9(0.013) (2.0, 3.0, 4.5)	8.2(0.053) (3.0, 7.0, 17.0)	20.95%	31.95%	29.05%
(100, 500)	Lasso	(0.045, 0.000)	36	51	60	27	15	27	2.4(0.011) (1.0, 2.0, 4.0)	8.8(0.022) (5.5, 9.0, 12.0)	12.53%	35.57%	34.80%
	w/o LS	(0.036, 0.028)	81	81	66	57	3	12	3.3(0.011) (2.0, 4.0, 4.5)	8.9(0.052) (4.0, 8.0, 13.5)	18.57%	30.90%	31.73%
	w/ LS	(0.045, 0.037)	45	60	60	39	6	27	2.6(0.011) (1.0, 2.0, 4.0)	10.1(0.030) (6.5, 10.0, 13.0)	13.13%	34.97%	35.17%
(200, 500)	Lasso	(0.043, 0.000)	44	94	76	46	22	10	2.9(0.011) (2.0, 3.0, 4.0)	4.2(0.020) (2.0, 4.0, 6.5)	21.69%	25.82%	25.07%
	w/o LS	(0.035, 0.010)	66	96	90	58	34	14	3.6(0.011) (2.0, 4.0, 5.0)	7.4(0.048) (3.0, 6.0, 12.0)	20.27%	27.18%	26.26%
	w/ LS	(0.043, 0.026)	64	92	74	60	16	4	3.1(0.010) (2.0, 3.0, 4.0)	5.6(0.028) (2.0, 5.0, 9.5)	21.87%	26.66%	25.38%

Table 10: Simulation results for $\rho = 0.9$ based on BIC

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.037, 0.000)	40	70	70	25	20	50	2.8(0.011) (1.5, 2.5, 4.5)	5.2(0.029) (3.0, 5.0, 9.0)	18.95%	NA	28.10%
	w/o LS	(0.036, 0.001)	45	70	70	25	20	50	2.8(0.011) (1.5, 3.0, 4.5)	5.7(0.028) (3.0, 5.0, 10.0)	18.85%	NA	28.60%
	w/ LS	(0.036, 0.008)	50	75	65	30	30	45	3.0(0.013) (1.5, 3.0, 5.0)	7.8(0.075) (3.0, 5.0, 21.0)	17.15%	NA	28.95%
(100, 500)	Lasso	(0.038, 0.000)	36	54	60	36	18	36	2.7(0.010) (1.5, 3.0, 4.0)	15.3(0.047) (9.0, 15.0, 22.5)	4.30%	NA	38.07%
	w/o LS	(0.034, 0.002)	39	60	60	36	18	36	2.8(0.010) (2.0, 3.0, 4.5)	19.1(0.054) (10.5, 19.5, 25.0)	1.67%	NA	39.03%
	w/ LS	(0.038, 0.018)	42	66	66	39	18	33	2.9(0.011) (1.5, 3.0, 4.0)	16.8(0.042) (9.5, 17.0, 21.0)	3.90%	NA	37.40%
(200, 500)	Lasso	(0.028, 0.000)	60	94	84	60	52	44	3.9(0.012) (2.5, 4.0, 6.0)	19.3(0.114) (7.0, 14.5, 36.0)	10.20%	NA	30.87%
	w/o LS	(0.028, 0.001)	58	94	84	58	52	46	3.9(0.012) (2.0, 4.0, 5.5)	18.2(0.124) (6.5, 14.5, 34.0)	11.93%	NA	30.58%
	w/ LS	(0.028, 0.013)	62	98	92	60	58	40	4.1(0.012) (2.5, 4.0, 6.0)	22.2(0.134) (6.0, 20.0, 40.5)	9.45%	NA	30.97%

Table 11: Simulation results for $\rho = 0.9$ based on BICC

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.042, 0.000)	40	70	65	25	15	35	2.5(0.011) (1.0, 2.0, 4.5)	5.2(0.019) (3.0, 5.0, 7.5)	20.50%	NA	27.25%
	w/o LS	(0.042, 0.008)	45	70	65	30	15	35	2.6(0.011) (1.5, 2.0, 4.5)	8.1(0.051) (4.0, 6.5, 16.0)	20.55%	NA	27.20%
	w/ LS	(0.042, 0.005)	45	70	65	25	15	35	2.5(0.011) (1.0, 2.0, 4.5)	8.2(0.052) (3.0, 7.0, 17.0)	20.40%	NA	27.50%
(100, 500)	Lasso	(0.046, 0.000)	36	51	57	27	15	24	2.3(0.011) (1.0, 2.0, 4.0)	8.8(0.022) (5.5, 9.0, 12.0)	12.63%	NA	34.70%
	w/o LS	(0.045, 0.026)	45	51	45	24	3	18	2.1(0.009) (1.0, 2.0, 3.0)	8.9(0.052) (4.0, 8.0, 13.5)	18.63%	NA	31.23%
	w/ LS	(0.046, 0.010)	42	54	54	30	12	24	2.4(0.011) (1.0, 2.0, 4.0)	10.1(0.030) (6.5, 10.0, 13.0)	12.63%	NA	35.40%
(200, 500)	Lasso	(0.043, 0.000)	40	94	72	42	16	4	2.7(0.009) (2.0, 3.0, 4.0)	4.2(0.020) (2.0, 4.0, 6.5)	21.69%	NA	25.65%
	w/o LS	(0.043, 0.001)	42	92	70	46	16	8	2.7(0.010) (1.5, 3.0, 4.0)	7.4(0.048) (3.0, 6.0, 12.0)	21.47%	NA	25.66%
	w/ LS	(0.043, 0.002)	44	94	72	46	18	8	2.8(0.010) (2.0, 3.0, 4.0)	5.6(0.028) (2.0, 5.0, 9.5)	21.48%	NA	25.58%

Table 12: Simulation results for $\rho = 0.9$ based on EBIC

(n, p)	Distance	$(\bar{\lambda}_{1,\text{opt}}, \bar{\lambda}_{2,\text{opt}})$	β_1	β_2	β_7	β_9	β_{21}	β_{22}	N_{true}	$N_{\text{non-zero}}$	Training error	Tuning error	Testing error
(100, 50)	Lasso	(0.043, 0.000)	40	65	65	25	15	30	2.4(0.011) (1.0, 2.0, 4.0)	3.9(0.013) (2.0, 4.0, 5.5)	21.00%	NA	27.30%
	w/o LS	(0.044, 0.033)	50	75	45	25	0	15	2.1(0.008) (1.0, 2.0, 3.0)	3.8(0.018) (1.5, 4.0, 5.5)	24.25%	NA	26.65%
	w/ LS	(0.042, 0.025)	65	80	60	25	15	25	2.7(0.010) (1.5, 3.0, 4.0)	5.3(0.032) (2.5, 5.0, 7.0)	20.30%	NA	27.90%
(100, 500)	Lasso	(0.046, 0.000)	36	51	57	27	15	24	2.3(0.011) (1.0, 2.0, 4.0)	8.7(0.022) (5.5, 9.0, 11.5)	12.63%	NA	34.70%
	w/o LS	(0.045, 0.042)	57	57	27	27	0	3	1.9(0.008) (1.0, 2.0, 2.5)	4.3(0.021) (2.0, 4.0, 8.0)	24.20%	NA	29.20%
	w/ LS	(0.046, 0.012)	42	54	54	30	12	24	2.4(0.011) (1.0, 2.0, 4.0)	9.0(0.023) (5.5, 9.0, 12.0)	12.93%	NA	35.13%
(200, 500)	Lasso	(0.045, 0.000)	40	92	64	40	12	4	2.5(0.009) (1.0, 3.0, 4.0)	3.4(0.012) (2.0, 3.0, 5.0)	22.34%	NA	25.08%
	w/o LS	(0.045, 0.021)	48	90	44	30	12	4	2.3(0.008) (1.0, 2.0, 3.0)	3.4(0.012) (2.0, 3.0, 5.0)	23.93%	NA	26.39%
	w/ LS	(0.044, 0.004)	40	92	64	44	12	4	2.6(0.010) (1.0, 3.0, 4.0)	3.6(0.014) (2.0, 3.0, 5.5)	22.29%	NA	25.39%

Table 13: Selected predictors in chromosome 8 for coeliac disease

SNP	BP position	P-Value	Gene	Function
gender		0.00000		
rs6995469	3510038	0.00010	CSMD1	Potential suppressor of squamous cell carcinomas
rs6991080	4612685	0.02649	CSMD1	Potential suppressor of squamous cell carcinomas
rs2924750	4902851	0.00228	intergenic	
rs10503561	15843279	0.00361		
rs13265570	20230308	0.00045	intergenic	
rs3943520	22338677	0.02305	SLC39A14	May be able to transport iron (by similarity) and acts as a zinc-influx transporter
rs1485750	25864862	0.00036	EBF2	Belongs to the conserved Olf/EBF family of helix-loop-helix transcription factors
rs2347501	32413693	0.00191	NRG1	Important for the normal development of the nervous system and the heart
rs2317630	57748410	0.00097	intergenic	
rs10106681	58428893	0.00243	intergenic	
rs10504244	59147328	0.01308	C8orf72	May be involved in tumor progression
rs10808759	69314054	0.00394	intergenic	
rs2726312	71622060	0.00137	intergenic	
rs3919902	72184730	0.01504	intergenic	
rs1445401	72411576	0.00200	EYA1	Encodes a member of the eyes absent (EYA) family of proteins. The encoded protein may play a role in the developing kidney, branchial arches, eye and ear
rs2977330	76857217	0.00044	intergenic	
rs894153	96301923	0.00144	intergenic	
rs736191	99264380	0.00159	intergenic	
rs648119	103207221	0.00005	NCALD	May be involved in the calcium-dependent regulation of rhodopsin phosphorylation. Binds three calcium ions
rs10096287	104864059	0.00064	RIMS2	Rab effector involved in exocytosis. May act as scaffold protein
rs4559272	119595143	0.00136	SAMD12	
rs4871072	121662368	0.00162	SNTB1	Adapter protein that binds to and probably organizes the subcellular localization of a variety of membrane proteins. May link various receptors to the actin cytoskeleton and the dystrophin glycoprotein complex
rs7386962	124149521	0.00067		
rs2122835	128347495	0.00065	intergenic	
rs10505604	134096770	0.00002	TG	A thyroid hormone precursor, storage of iodine, and storage of inactive thyroid hormones.
rs2896714	134564796	0.00204	ST3GAL1	Complex and appears to produce several proteins with no sequence overlap
rs7460819	136804688	0.00254	intergenic	
rs1499447	138051471	0.00000	intergenic	

References

- Ahn, A. and Kunkel, L. (1995), “Syntrophin binds to an alternatively spliced exon of dystrophin.” *The Journal of Cell Biology*, 128, 363–371.
- Alexanian, A., Bamberg, J., Hidaka, H., and Mornet, D. (2001), “Calcium-dependent regulation of interactions of caldesmon with calcium-binding proteins found in growth cones of chick forebrain neurons,” *Cellular and Molecular Neurobiology*, 21, 437–451.
- Bertsekas, D., Hager, W., and Mangasarian, O. (1999), *Nonlinear Programming*, Athena Scientific: Belmont, MA.
- Boat, T., Welsh, M., Beaudet, A., Scriver, C., Beaudet, A., Sly, W., and Valle, D. (1989), *The Metabolic Basis of Inherited Disease*, Mc Graw-Hill New York.
- Brookes, A. (1999), “The essence of SNPs,” *Gene*, 234, 177–186.
- Cantor, R., Lange, K., and Sinsheimer, J. (2010), “Prioritizing GWAS results: a review of statistical methods and recommendations for their application,” *The American Journal of Human Genetics*, 86, 6–22.
- CDC (2011), “Tobacco Use, targeting the nations leading killer at a glance 2011,” <http://www.cdc.gov/chronicdisease/resources/publications/AAG/osh.htm>.
- Chen, J. and Chen, Z. (2008), “Extended Bayesian information criteria for model selection with large model spaces,” *Biometrika*, 95, 759–771.
- Chen, S., Donoho, D., and Saunders, M. (1999), “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, 20, 33–61.
- Claerbout, J. and Muir, F. (1973), “Robust modeling with erratic data,” *Geophysics*, 38, 826–844.
- Cook, P., Ju, B., Telese, F., Wang, X., Glass, C., and Rosenfeld, M. (2009), “Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions,” *Nature*, 458, 591–596.
- Daubechies, I., Defrise, M., and De Mol, C. (2004), “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Mathematics*, 57, 1413–1457.
- Dauchet, L., Amouyel, P., Hercberg, S., and Dallongeville, J. (2006), “Fruit and vegetable consumption and risk of coronary heart disease: a meta-analysis of cohort studies,” *The Journal of Nutrition*, 136, 2588–2593.
- Daye, Z. and Jeng, X. (2009), “Shrinkage and model selection with correlated variables via weighted fusion,” *Computational Statistics & Data Analysis*, 53, 1284–1298.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), “Least angle regression,” *The Annals of Statistics*, 32, 407–499.
- 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.

- Fasano, A., Berti, I., Gerarduzzi, T., Colletti, R., Drago, S., Elitsur, Y., Green, P., Guandalini, S., Hill, I., Pietzak, M., et al. (2003), “Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study,” *Archives of Internal Medicine*, 163, 286–292.
- Ford, D., Easton, D., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D., Weber, B., Lenoir, G., Chang-Claude, J., et al. (1998), “Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families,” *The American Journal of Human Genetics*, 62, 676–689.
- Ford, E. and Mokdad, A. (2001), “Fruit and vegetable consumption and diabetes mellitus incidence among US adults,” *Preventive Medicine*, 32, 33–39.
- Friedman, J., Hastie, T., Höfling, H., and Tibshirani, R. (2007), “Pathwise coordinate optimization,” *The Annals of Applied Statistics*, 1, 302–332.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010), “Regularization paths for generalized linear models via coordinate descent,” *Journal of Statistical Software*, 33, 1–22.
- Fu, W. (1998), “Penalized regressions: the bridge versus the lasso,” *Journal of Computational and Graphical Statistics*, 7, 397–416.
- Goldstein, D. (2009), “Common genetic variation and human traits,” *New England Journal of Medicine*, 360, 1696–1698.
- Hauge, H., Patzke, S., and Aasheim, H. (2007), “Characterization of the FAM110 gene family,” *Genomics*, 90, 14–27.
- Håvik, B., Le Hellard, S., Rietschel, M., Lybæk, H., Djurovic, S., Mattheisen, M., Mühleisen, T., Degenhardt, F., Priebe, L., and Maier, W. (2011), “The complement control-related genes CSMD1 and CSMD2 associate to schizophrenia,” *Biological Psychiatry*, 70, 35–42.
- He, F., Nowson, C., and MacGregor, G. (2006), “Fruit and vegetable consumption and stroke: meta-analysis of cohort studies,” *The Lancet*, 367, 320–326.
- Huang, H., Chin, S., Ginestier, C., Bardou, V., Adélaïde, J., Iyer, N., Garcia, M., Pole, J., Callagy, G., Hewitt, S., et al. (2004), “A recurrent chromosome breakpoint in breast cancer at the NRG1/neuregulin 1/hereregulin gene,” *Cancer Research*, 64, 6840–6844.
- Kim, J., Kim, Y., and Kim, Y. (2008), “A gradient-based optimization algorithm for lasso,” *Journal of Computational and Graphical Statistics*, 17, 994–1009.
- Kim, Y. and Kim, J. (2004), “Gradient LASSO for feature selection,” in *Proceedings of the twenty-first international conference on Machine learning*, ACM, pp. 60–67.
- Kitagawa, H. and Paulson, J. (1994), “Differential expression of five sialyltransferase genes in human tissues,” *Journal of Biological Chemistry*, 269, 17872–17878.
- Knight, K. and Fu, W. (2000), “Asymptotics for lasso-type estimators,” *The Annals of Statistics*, 28, 1356–1378.

- Koh, K., Kim, S., and Boyd, S. (2007), “An interior-point method for large-scale l_1 -regularized logistic regression,” *Journal of Machine Learning Research*, 8, 1519–1555.
- Land, S. and Friedman, J. (1997), “Variable fusion: A new adaptive signal regression method,” Tech. rep., Department of Statistics, Carnegie Mellon University, Pittsburg, PA.
- Lau, W. and Scholnick, S. (2003), “Identification of two new members of the CSMD gene family,” *Genomics*, 82, 412–415.
- Lee, S., Lee, H., Abbeel, P., and Ng, A. (2006), “Efficient L_1 Regularized Logistic Regression,” in *Proceedings of the National Conference on Artificial Intelligence*, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, vol. 21, pp. 401–408.
- Liu, J., Yuan, L., and Ye, J. (2010), “An efficient algorithm for a class of fused lasso problems,” *Prostate*, 98, 98–107.
- Losowsky, M. (2008), “A history of coeliac disease,” *Digestive Diseases*, 26, 112–120.
- Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L., Ding, W., et al. (1994), “A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1,” *Science*, 266, 66–71.
- Nesterov, Y. (1983), “A method of solving a convex programming problem with convergence rate $O(1/2^k)$,” in *Soviet Mathematics Doklady*, vol. 27, pp. 372–376.
- (2003), *Introductory lectures on convex optimization: a basic course*, Kluwer Academic Publishers.
- (2007), “Gradient methods for minimizing composite objective function,” *CORE Report*, available at http://www.ecore.be/DPs/dp_1191313936.pdf.
- NIH (2011), “What are single nucleotide polymorphisms (SNPs)?” <http://ghr.nlm.nih.gov/handbook/genomicresearch/snp/>.
- Osborne, M., Presnell, B., and Turlach, B. (2000), “On the LASSO and its dual,” *Journal of Computational and Graphical statistics*, 9, 319–337.
- Park, M. and Hastie, T. (2007), “ L_1 -regularization path algorithm for generalized linear models,” *Journal of the Royal Statistical Society Series B*, 69, 659–677.
- (2008), “Penalized logistic regression for detecting gene interactions,” *Biostatistics*, 9, 30–50.
- Peles, E., Bacus, S., Koski, R., Lu, H., Wen, D., Ogden, S., Levy, R., and Yarden, Y. (1992), “Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein that induces differentiation of mammary tumor cells,” *Cell*, 69, 205–216.
- Plowman, G., Green, J., Culouscou, J., Carlton, G., Rothwell, V., and Buckley, S. (1993), “Heregulin induces tyrosine phosphorylation of HER4/p180erbB4,” 366, 473–475.
- Rewers, M. (2005), “Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease?” *Gastroenterology*, S47–S51.

- Rinaldo, A. (2009), “Properties and refinements of the fused lasso,” *The Annals of Statistics*, 37, 2922–2952.
- Rodrigo, L. (2006), “Celiac disease,” *World Journal of Gastroenterology*, 6585–6593.
- Rose, J., Behm, F., Drgon, T., Johnson, C., and Uhl, G. (2010), “Personalized smoking cessation: interactions between nicotine dose, dependence and quit-success genotype score,” *Molecular Medicine*, 16, 247–253.
- Santosa, F. and Symes, W. (1986), “Linear inversion of band-limited reflection seismograms,” *SIAM Journal on Scientific and Statistical Computing*, 7, 1307–1330.
- Sasco, A., Secretan, M., and Straif, K. (2004), “Tobacco smoking and cancer: a brief review of recent epidemiological evidence,” *Lung Cancer*, 45, S3–S9.
- Schwarz, G. (1978), “Estimating the dimension of a model,” *The Annals of Statistics*, 6, 461–464.
- Sha, F., Park, Y., and Saul, L. (2007), “Multiplicative Updates for L1-Regularized Linear and Logistic Regression,” *Advances in Intelligent Data Analysis VII*, 13–24.
- Steinmetz, K. and Potter, J. (1996), “Vegetables, fruit, and cancer prevention: a review,” *Journal of the American Dietetic Association*, 96, 1027–1039.
- Sullivan, A., Wigginton, J., and Kirschner, D. (2001), “The coreceptor mutation CCR5 Δ 32 influences the dynamics of HIV epidemics and is selected for by HIV,” *Proceedings of the National Academy of Sciences*, 98, 10214–10219.
- Taylor, H., Banks, S., and McCoy, J. (1979), “Deconvolution with the l1 norm,” *Geophysics*, 44, 39–52.
- Taylor, K., Morgan, H., Johnson, A., and Nicholson, R. (2005), “Structure–function analysis of a novel member of the LIV-1 subfamily of zinc transporters, ZIP14,” *FEBS Letters*, 579, 427–432.
- Tibshirani, R. (1996), “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society Series B*, 267–288.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005), “Sparsity and smoothness via the fused lasso,” *Journal of the Royal Statistical Society Series B*, 67, 91–108.
- Trynka, G., Zhernakova, A., Romanos, J., Franke, L., Hunt, K., Turner, G., Bruinenberg, M., Heap, G., Platteel, M., Ryan, A., et al. (2009), “Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF- κ B signalling,” *Gut*, 58, 1078–1083.
- van Heel, D., Franke, L., Hunt, K., Gwilliam, R., Zhernakova, A., Inouye, M., Wapenaar, M., Barnardo, M., Bethel, G., Holmes, G., et al. (2007), “A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21,” *Nature Genetics*, 39, 827–829.

- van Heel, D. and West, J. (2006), “Recent advances in coeliac disease,” *Gut*, 55, 1037–1046.
- Venkitaraman, A. (2002), “Cancer susceptibility and the functions of BRCA1 and BRCA2,” *Cell*, 108, 171–182.
- Vignal, A., Milan, D., SanCristobal, M., Eggen, A., et al. (2002), “A review on SNP and other types of molecular markers and their use in animal genetics,” *Genetics Selection Evolution*, 34, 275–306.
- Wang, S., Betz, A., and Reed, R. (2002), “Cloning of a novel Olf-1/EBF-like gene, O/E-4, by degenerate oligo-based direct selection,” *Molecular and Cellular Neuroscience*, 20, 404–414.
- Wang, Y. and Südhof, T. (2003), “Genomic definition of RIM proteins: evolutionary amplification of a family of synaptic regulatory proteins,” *Genomics*, 81, 126–137.
- Wolters, V. and Wijmenga, C. (2008), “Genetic background of celiac disease and its clinical implications,” *The American Journal of Gastroenterology*, 103, 190–195.
- Wu, T., Chen, Y., Hastie, T., Sobel, E., and Lange, K. (2009), “Genome-wide association analysis by lasso penalized logistic regression,” *Bioinformatics*, 25, 714–721.
- Wu, T. and Lange, K. (2008), “Coordinate descent algorithms for lasso penalized regression,” *The Annals of Applied Statistics*, 2, 224–244.
- Zandi, P., Zöllner, S., Avramopoulos, D., Willour, V., Chen, Y., Qin, Z., Burmeister, M., Miao, K., Gopalakrishnan, S., McEachin, R., et al. (2008), “Family-based SNP association study on 8q24 in bipolar disorder,” *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 147, 612–618.
- Zhang, P., Xiang, N., Chen, Y., Śliwerska, E., McInnis, M., Burmeister, M., and Zöllner, S. (2010), “Family-based association analysis to finemap bipolar linkage peak on chromosome 8q24 using 2,500 genotyped SNPs and 15,000 imputed SNPs,” *Bipolar Disorders*, 12, 786–792.
- Zhao, X., Shi, Y., Tang, J., Tang, R., Yu, L., Gu, N., Feng, G., Zhu, S., Liu, H., Xing, Y., et al. (2004), “A case control and family based association study of the neuregulin 1 gene and schizophrenia,” *Journal of Medical Genetics*, 41, 31–34.