Published in final edited form as:

Electron J Stat. 2016 ; 10(2): 2312–2328. doi:10.1214/16-EJS1169.

Designing penalty functions in high dimensional problems: The role of tuning parameters

Ting-Huei Chen,

Department of Mathematics and Statistics, Laval University, Quebec, QC G1V0A6, Canada

Wei Sun[*], and

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

Jason P. Fine

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Abstract

Various forms of penalty functions have been developed for regularized estimation and variable selection. Screening approaches are often used to reduce the number of covariate before penalized estimation. However, in certain problems, the number of covariates remains large after screening. For example, in genome-wide association (GWA) studies, the purpose is to identify Single Nucleotide Polymorphisms (SNPs) that are associated with certain traits, and typically there are millions of SNPs and thousands of samples. Because of the strong correlation of nearby SNPs, screening can only reduce the number of SNPs from millions to tens of thousands and the variable selection problem remains very challenging. Several penalty functions have been proposed for such high dimensional data. However, it is unclear which class of penalty functions is the appropriate choice for a particular application. In this paper, we conduct a theoretical analysis to relate the ranges of tuning parameters of various penalty functions with the dimensionality of the problem and the minimum effect size. We exemplify our theoretical results in several penalty functions. The results suggest that a class of penalty functions that bridges L_0 and L_1 penalties requires less restrictive conditions on dimensionality and minimum effect sizes in order to attain the two fundamental goals of penalized estimation: to penalize all the noise to be zero and to obtain unbiased estimation of the true signals. The penalties such as SICA and Log belong to this class, but they have not been used often in applications. The simulation and real data analysis using GWAS data suggest the promising applicability of such class of penalties.

Keywords and phrases

Folded-concave penalties; tuning parameter selection; genome-wide association studies

[*]This work was supported by in part by NIH grant R01GM105785.

Supplementary Material

Designing penalty functions in high dimensional problems: The role of tuning parameters – Supplementary Material (doi: 10.1214/16-EJS1169SUPP;.pdf).

1. Introduction

In genome-wide association (GWA) studies, the goal is to identify the genetic factors such as single nucleotide polymorphisms (SNPs) that are associated with diseases. With the availability of a dense map of SNPs, it is statistically very challenging to select the important SNPs from millions of SNPs using only a couple of thousand samples. Regularized estimation procedures can be applied for simultaneous selection of important variables (SNPs) and estimation of their effects for high dimensional data in GWA studies. The objective function of the regularized estimation is composed of a model fitting metric (e.g., likelihood function) and a penalty function for the parameters subject to regularization. Prior to the usage of regularized estimation, screening can be applied to reduce the number of SNPs to be considered for penalized estimation. However, due to the high correlation of neighboring SNPs, the number of SNPs that pass a reasonable screening criterion is often larger than or much larger than the sample size.

We use the real SNP genotype data from a recent study (Wright et al., 2014) to illustrate the correlation structure of genotype data. We take the genotypes of 645,316 SNPs in chromosome 1 from 1,198 samples, and randomly pick 30 SNPs as important variables to simulate the response under the linear model. The effect size is simulated as 0.7 and the residual errors are standard normal variables. Figure 1 shows a Manhattan plot of the marginal association p-values. The 30 important SNPs are labeled by grey vertical lines. It is obvious that the high correlation among nearby SNPs leads to small p-values for those SNPs which are close to the 30 important SNPs. If we apply screening using the p-value cut-off 10^{-4}, 3,087 SNPs will be selected which include 20 of the 30 important SNPs. Alternatively, if the p-value cut-off is 10^{-8}, 991 SNPs will be selected, which include only 13 of the 30 important SNPs. Thus screening method can be helpful to certain extend, and screening with stringent threshold would lead to many false negatives. This conclusion is consistent with the extensive empirical study by Bühlmann and Mandozzi (2012). Therefore, the penalty function itself is still the key for high dimensional data analysis, and it is desirable to identify penalty functions that can tolerate higher dimension.

Several penalty functions have been proposed for high dimensional data analysis. One of the most popular penalty functions is the Lasso penalty (Tibshirani, 1996). The variable selection consistency of the Lasso requires the irrepresentable condition (Zhao and Yu, 2006) that there is no strong correlation between the "*important covariates*" that have non-zero effects and the "*unimportant covariates*" that have zero effects. This condition may not be satisfied in some applications, such as GWA studies. Recent studies have shown that a class of folded concave penalties can achieve variable selection consistency without requiring such an irrepresentable condition (Fan and Lv, 2010). These folded concave penalties include, but are not limited to SCAD (Smoothly Clipped Absolute Deviation) (Fan, 1997; Fan and Li, 2001), MCP (Minimax Concave Penalty) (Zhang, 2010), SICA (Smooth Integration of Counting and Absolute deviation) (Lv and Fan, 2009), and a Log penalty (Friedman (2008), Sun, Ibrahim and Zou (2010)).

A common concern in real data applications of penalized estimation is to tune the regularization parameters to achieve the two fundamental goals of penalized estimation: to

penalize all the noise to be zero and to obtain an unbiased estimation of the true signals. However, it may not be clear whether such "optimal" tuning is possible, and this is the focus of our study. Mazumder, Friedman and Hastie (2011) study the non-convex optimization problem for SCAD, MCP and Log penalties, but they did not address the roles of tuning parameters of those penalties in variable selection. Moreover, all the aforementioned folded-concave penalties have two tuning parameters, and thus in practice, the immediate questions concern whether they both should be tuned, and what is the consequence of tuning only one of them in order to improve computational efficiency. Previous work has provided recommendations regarding the choice of tuning parameters, but there is no systematic asymptotic studies on the roles of multiple tuning parameters. To address those issues, we will relate the choice of tuning parameters to the difficulty of the variable selection problem, namely the minimum effect size and the dimensions, i.e., the number of important and unimportant covariates.

The results suggest that a class of penalty functions that bridges L_0 and L_1 penalties such as Log and SICA requires less restrictive conditions on dimensionality and minimum effect sizes, while achieving the two fundamental goals of penalized estimation. For the tuning of the regularization parameters, our study shows that both SICA and Log penalties have very limited performance if only one of the two regularization parameters is tuned, while tuning both regularization parameters can significantly improve their performances, although at the price of heavier computational burden. Our results are also insightful for designing other penalty functions. For example, our results imply that two tuning parameters are sufficient to achieve the two fundamental goals. Therefore, penalties with more than two regularization parameters may not be needed due to the substantial increase of computational cost.

We conducted empirical analyses of the penalty functions using both simulated data and real data in GWA settings. Those empirical results support the idea that the class of penalty functions that bridges L_0 and L_1 holds promise for genomic studies.

2. Theoretical results

2.1. Notations and problem setup

Let $p_{\varpi}(\beta)$ be a penalty function of β, where ϖ are regularization parameters with arbitrary dimension. $p_{\varpi}(\beta)$ is referred to as a folded concave penalty if it satisfies the following condition:

Condition 1—$p_{\varpi}(\beta)$ is concave in $\beta \in [0, \infty)$, with continuous derivative $p'_{\varpi}(\beta) \geq 0$, and $p'_{\varpi}(0+) > 0$.

We formulate the effects of the covariates via a generalized linear regression model, permitting continuous and discrete outcome variables. Consider a sample of n responses, $y = (y_1, \ldots, y_n)^\top$, where each y_i, $i = 1, \ldots, n$, is independently generated from an exponential family distribution with a density: $p(y_i|\theta_i) = \exp\{[y_i\theta_i - b(\theta_i)]/\phi + c(y_i, \phi)\}$, where θ_i is the canonical parameter and $\phi \in (0, \infty)$ is the dispersion parameter. Let x_{ij} be the value of the j-th covariate in the i-th sample, and let $X = (x_{ij})$ be a $n \times p$ matrix of the covariates' values.

We assume that X has been normalized such that $\sum_{i=1}^{n} x_{ij}^2 = n$, for $j = 1, \ldots, p$. Under the assumed generalized linear model, $\theta_i = \sum_{j=1}^{p} x_{ij}\beta_j$, where β_j's are regression coefficients. Let $E(y) = \mu(\theta) = (\partial_{\theta_1} b(\theta_1), \ldots, \partial_{\theta_n} b(\theta_n))^\top$ and $\sum(\theta) = \mathrm{diag}\{\partial_{\theta_1}^2 b(\theta_1), \ldots, \partial_{\theta_n}^2 b(\theta_n)\}$. We maximize the penalized likelihood $Q_n(\beta) = l_n(\beta) - \sum_{j=1}^{p} p_\varpi(|\beta_j|)$, where $l_n(\beta) = n^{-1}[y^\top\theta - \mathbf{1}^\top b(\theta)]$ is an affine transformation of the log-likelihood.

Without loss of generality, we assume that the first s covariates of X are important (i.e., having non-zero effect on the response variable) and denote them collectively by X_1, and then denote the remaining $p-s$ unimportant covariates by X_2, such that $X = (X_1, X_2)$. Similarly, we partition β for the important and unimportant covariates such that $\beta = (\beta_1^\top, \beta_2^\top)^\top$. Let $\beta_0 = (\beta_{01}^\top, \beta_{02}^\top)^\top = (\beta_{01}, \ldots, \beta_{0p})^\top$ be the true coefficients, such that $\beta_{02} = 0$. Let θ_0 be the true values of θ such that $\theta_0 = X\beta_0$.

It is difficult to analytically study the global maximizer of the penalized likelihood. Following the previous work (Fan and Lv, 2011), we study the local maximizer of the penalized likelihood that satisfies a set of sufficient and almost necessary conditions specified in Theorem 1 (see Appendix).

2.2. The role of the tuning parameters

The dimension of the regression problem and the minimum effect size are assumed to satisfy the following conditions:

Condition 2.1—$\log p = O(n^\alpha)$ and $s = O(n^\nu)$, respectively, with $0 \le \alpha < 1$ and $0 \le \nu < 1/2$.

Condition 2.2—$d_n \equiv 2^{-1} \min_{1 \le j \le s}\{|\beta_{j0}|\} = O(n^{-\gamma_0}(\log n)^{1/2})$ for some $\gamma_0 \in (\nu, 1/2)$.

The restriction of $\gamma_0 > \nu$ (which is equivalent to $s < n^{\gamma_0}$) in Condition 2.2 can be understood as an identifiability condition so that $d_n s = O(n^{\nu-\gamma_0}(\log n)^{1/2})$ can be bounded by a constant. Otherwise the response variable is unbounded, with non-trivial probability.

A maximizer of the penalized likelihood, $\hat\beta = (\hat\beta_1^\top, \hat\beta_2^\top)^\top$, is considered to have weak oracle property if $\hat\beta_2 = 0$ with probability tending to 1 as $n \to \infty$, and $\hat\beta_1$ is consistent under L_∞ loss (Lv and Fan, 2009). We will study the role of tuning parameters by studying the conditions for the weak oracle property. To this end, we generalize the conditions for the weak oracle property in Fan and Lv (2011) to impose constraints on the penalty function rather than particular tuning parameters, which gives the following Conditions 3.1–3.3. This generalization is necessary because the original conditions are too stringent for any penalty function whose $p'_\varpi(0+)$ involves more than one tuning parameter. For example, the Log penalty cannot satisfy the original conditions for the weak oracle property. After generalizing the conditions, we can show that the Log penalty can indeed fulfill the conditions of the weak oracle property.

Condition 3.1— $p'_{\varpi}(d_n) \ll b_s^{-1} d_n$, where $b_s \equiv O(n^{\gamma_s}) = O(n \| [X_1^\top \sum(\theta_0) X_1]^{-1} \|_\infty)$ with $\gamma_s \geq 0$. A corollary of Condition 3.1 is $p'_{\varpi}(d_n) \ll d_n$.

Condition 3.2—

$$\| X_2^\top \sum(\theta_0) X_1 [X_1^\top \sum(\theta_0) X_1]^{-1} \|_\infty \leq \min \{ K p'_{\varpi}(0+) / p'_{\varpi}(d_n), O(n^\nu) \}$$

for $K \in (0, 1)$.

Condition 3.3— $p'_{\varpi}(0+) \gg \max(n^{-2\gamma_0 + 2\nu} \log n, n^{\nu - 1/2} (\log n)^{1/2})$ and $p'_{\varpi}(0+) > \eta_p \sigma^{-1/2} (1-K)^{-1}$, where K is defined in Condition 3.2, σ is a constant that is defined based on the range of the response variable y (see proposition A1 in the Supplementary Materials for details), and $\eta_p = n^{-1/2+a/2} (\log n)^{1/2}$.

Condition 3.1 requires the derivative of the penalty function (i.e., the increase of penalization as the regression coefficient increases) for important covariates to be small enough. Condition 3.2 says that the ratio of the penalties' derivatives for unimportant covariates and for important ones $(p'_{\varpi}(0+)/p'_{\varpi}(d_n))$ should be large enough relative to the maximum correlation between important and unimportant covariates, which is a generalization of the irrepresentable condition for Lasso (Zhao and Yu, 2006). Condition 3.3 requires the derivative of the penalty function for unimportant covariates to be large enough. In contrast to the conditions for the weak oracle property in Fan and Lv (2011), a critical modification is that we restrict the size of $p'_{\varpi}(0+)$ in Condition 3.3, which replaces the condition $\lambda_n \gg n^{-a} (\log n)^2$ stated in equation (18) of Fan and Lv (2011). For SCAD and MCP, $p'_{\varpi}(0+) = \lambda_n$, and thus constraints on λ_n or $p'_{\varpi}(0+)$ are equivalent. However, for Log and SICA, $p'_{\varpi}(0+) = O(\lambda_n / \tau_n)$. Therefore, the generalized condition only requires the ratio of the two regularization parameters to be large enough instead of imposing a constraint on λ_n itself. Given Conditions 2.1–2.2, Conditions 3.1–3.3, and Conditions 4.1–4.4 (presented in the Appendix), which are for the design matrix X, we have the weak oracle property (Theorem 2 in the Appendix).

One immediate conclusion from Conditions 3.1–3.3 is that the constraints on the penalty function $p_{\lambda}(\beta)$ are applied on the two quantities $p'_{\varpi}(0+)$ and $p'_{\varpi}(d_n)$. With an appropriate design, two tuning parameters can give enough degrees of freedom on these two quantities so that Conditions 3.1–3.3 are satisfied.

Next we discuss the implications of Conditions 3.1–3.3 for the four folded concave penalties: SCAD, MCP, Log, and SICA. It is more convenient to define SCAD and MCP by their derivatives.

$$p'_{\mathrm{SCAD}}(|\beta_j|; \lambda, a) = \{ \lambda I(|\beta_j| \leq \lambda) + [(a\lambda - |\beta_j|)/(a-1)] I(\lambda < |\beta_j| < a\lambda) \},$$

where $\lambda > 0$ and $a > 2$ are two regularization parameters.

$$p'_{\mathrm{MCP}}(|\beta_j|;\lambda,a)=I(|\beta_j|<a\lambda)(a\lambda-|\beta_j|)/a,$$

where $\lambda > 0$ and $a > 0$ are two regularization parameters. The Log and SICA penalties are defined as

$$p_{\log;\lambda,\tau}(|\beta_j|)=\lambda\log(|\beta_j|+\tau), \text{ and}$$
$$p_{\mathrm{SICA}}(|\beta_j|;\lambda,\tau)=\lambda\{I(|\beta_j|\neq 0)|\beta_j|/(|\beta_j|+\tau)+\tau|\beta_j|/(|\beta_j|+\tau)\},$$

respectively, where $\lambda > 0$ and $\tau > 0$ are two regularization parameters. In the following discussions, the tuning parameters employed by a penalty are indicated by subscripts. For example, the SCAD penalty with one tuning parameter λ_n (the other regularization parameter a being set as constant) is denoted by $\mathrm{SCAD}_{\lambda_n}$ and the SCAD penalty with two tuning parameters λ_n and a_n is denoted by $\mathrm{SCAD}_{\lambda_n,a_n}$.

Let $\eta_p = n^{-1/2+\alpha/2}(\log n)^{1/2}$, which is a monotone transformation of dimension $\log(p) = O(n^{\alpha})$. Let $\eta_d = \min(n^{\gamma_0/2}(\log n)^{-1/4}, n^{-\gamma_0+1/2})$, which, by Condition 2.2, is a function of the minimum effect size: $d_n \equiv \min_{1\le j\le s}\{|\beta_{j0}|\} = O(n^{-\gamma_0}(\log n)^{1/2})$. In the following propositions, we will discuss the properties of different penalties with respect to s (the number of non-zero coefficients), d_n, η_d, and η_p.

Proposition 1 ($\mathrm{SCAD}_{\lambda_n}$, $\mathrm{SCAD}_{\lambda_n,a_n}$, or MCP_{λ_n}). *If $d_n \gg \eta_p$ and $s \ll \eta_d$, there exist λ_n such that $d_n \gg \lambda_n > \eta_p$ to satisfy Conditions 3.1–3.3 for the weak oracle property. However, there is no such tuning parameter if $d_n \ll \eta_p$.*

Proposition 2 ($\mathrm{MCP}_{\lambda_n,a_n}$). *There are tuning parameters that satisfy Conditions 3.1–3.3 for the weak oracle property without further constraints other than $s \ll n^{\gamma_0}$, as is specified in Condition 2.2.*

Proposition 3 ($\mathrm{SICA}_{\lambda_n}$ or Log_{λ_n}). *There are tuning parameters that satisfy Conditions 3.1–3.3 for the weak oracle property if $d_n \gg \eta_p$, $s \ll \eta_d$, and*

$$\|X_2^{\top}\sum(\theta_0)X_1(X_1^{\top}\sum(\theta_0)X_1)^{-1}\|_{\infty} \le K\,(d_n/\tau+1)^2,$$

where $K \in (0, 1)$ was defined in Condition 3.3. There is no such tuning parameter if $d_n \ll \eta_p$.

Proposition 4 ($\mathrm{SICA}_{\lambda_n,\tau_n}$ or $\mathrm{Log}_{\lambda_n,\tau_n}$). *There are tuning parameters that satisfy Conditions 3.1–3.3 for the weak oracle property without further constraints other than $s \ll n^{\gamma_0}$, as is specified in Condition 2.2.*

Corollary 1 (Restriction on tuning parameter if $d_n \ll \eta_p$). *To satisfy Condition 3.1–3.3 requires* $a_n \to 0+$ *for* MCP_{λ_n, a_n}, *and* $\tau_n \to 0+$ *for* $SICA_{\lambda_n, \tau_n}$ *and* Log_{λ_n, τ_n}.

The proofs of Propositions 1–4 and Corollary 1 are presented in the Supplementary Materials (Chen et al., 2016).

By Proposition 1, if $d_n \gg \eta_p$ or $d_n \ll \eta_p$, SCAD has similar theoretical properties when one or two tuning parameters are used. This conclusion is consistent with many previous works where SCAD has satisfactory performance when the regularization parameter a is set to be a constant, e.g., 3.7. Using two tuning parameters (λ_n and a_n) does have some advantage over one tuning parameter (λ_n) when $d_n = O(\eta_p)$. However, since the situation of $d_n = O(\eta_p)$ only covers a negligible part of the space for d_n, we do not discuss it further here. Proposition 1 also states that if $d_n \ll \eta_p$ (the effect size is not large enough relative to the dimension), then there is no tuning parameter of SCAD to satisfy Conditions 3.1–3.3. Specifically, Condition 3.1 requires $p'_\varpi(d_n) \ll d_n$, and Condition 3.3 requires $p'_\varpi(0+) > c\eta_p$, where c is a constant. These two conditions cannot both be satisfied if $d_n \ll \eta_p$. Specifically, if SCAD satisfies Condition 3.3, then $p'_\varpi(0+) = \lambda_n > c\eta_p$. Given $d_n \ll \eta_p$ and $\eta_p < \lambda_n/c$, we have $d_n \ll \lambda_n$, and then we can show that $p'_\varpi(d_n) = \lambda_n$, which contradicts Condition 3.1. In addition, we can see that in this situation, both $p'_\varpi(0+)$ and $p'_\varpi(d_n)$ are functions of λ_n so that a plays no role in fulfilling Conditions 3.1 and 3.3. Therefore, tuning only one regularization parameter is sufficient and can be a computational advantage of SCAD.

By Propositions 1 and 2, tuning both λ_n and a_n significantly improves the performance of MCP if $d_n \ll \eta_p$. Specifically, if MCP satisfies Condition 3.3, then $p'_\varpi(0+) = \lambda_n > c\eta_p$. Then given $d_n \ll \eta_p$, we have $d_n \ll \lambda_n$. However, given a properly tuned $a_n = o(1)$ such that $d_n \geq a_n\lambda_n$, we have $p'_\varpi(d_n) = 0$, which allows MCP to satisfy Condition 3.1.

By Proposition 3, if we set $\tau = O(1)$ and only tune the regularization parameter λ, then $SICA_{\lambda_n}$ and Log_{λ_n} require the following condition to achieve the weak oracle property:

$$\|X_2^\top \sum(\theta_0) X_1 (X_1^\top \sum(\theta_0) X_1)^{-1}\|_\infty \leq K\,(d_n/\tau + 1).$$

This condition is similar to the irrepresentable condition of Lasso because when $\tau = O(1)$, $d_n/\tau + 1 \to 1$. Therefore, asymptotically $SICA_{\lambda_n}$ and Log_{λ_n} would perform in a way similar to Lasso. If $d_n \ll \eta_p$, then $SICA_{\lambda_n}$ and Log_{λ_n} cannot simultaneously satisfy Conditions 3.1 and 3.3, even if the irrepresentable condition is satisfied.

By Proposition 4, tuning both λ_n and τ_n significantly improves the performance of SICA and Log. Specifically, SICA and Log can have satisfactory variable selection performances even if the minimum effect size is much smaller with respect to the dimension of the problem: $d_n \ll \eta_p$. This can be justified by the following arguments. For Log penalty, $p'_\varpi(d_n) = p'_\varpi(0+)/(d_n/\tau_n + 1)$. Even Condition 3.3 requires a large value of $p'_\varpi(0+)$; a small enough τ_n can help $p'_\varpi(d_n)$ to satisfy Condition 3.1. SICA has similar properties since it has

$p'_{\varpi}(d_n)=p'_{\varpi}(0+)/(d_n/\tau_n+1)^2$. Therefore, the implications of Proposition 3 and Proposition 4 for the practical use of SICA and Log penalties would be that we should not treat τ as a constant.

Corollary 1 shows that for a difficult variable selection problem where $d_n \ll \eta_p$, the tuning parameter a_n of MCP or τ_n of SICA or Log should be on the scale of $o(1)$. Zhang (2010) suggests that a larger tuning parameter a in MCP leads to a bigger bias and less accurate variable selection, $a = 1$ leads to a singularity problem, and $a < 1$ leads to a dramatic increase in computational cost. Similarly, Lv and Fan (2009) suggest that for penalized estimates using SICA, the bias decreases to 0 as τ_n goes to 0+, but the computational difficulty increases because the maximum concavity goes to infinity. Similar conclusions apply to Log penalty. Although $\text{MCP}_{\lambda_n,a_n}$, $\text{SICA}_{\lambda_n,\tau_n}$, and $\text{Log}_{\lambda_n,\tau_n}$ have similar theoretical properties by Propositions 2 and 4, the following numerical studies show that the computation cost for SICA and Log is more affordable than that of MCP.

3. Algorithm and tuning parameter selection

We obtain the penalized estimates using SCAD or MCP by the coordinate descent algorithms implemented in the R package ncvreg (Breheny and Huang, 2011). We implement the penalized estimation using SICA and Log penalties by a combination of the coordinate descent algorithm and Local Linear Approximation (LLA) (Zou and Li, 2008). Specifically, we update the estimate of each regression coefficient sequentially (which is the coordinate decent part), and the solution of each coefficient is obtained after applying a local linear approximation. The details can be found in the Supplementary Materials.

We select a particular combination of tuning parameters from the initial tuning parameter pool using the extended BIC (Chen and Chen, 2008, 2012). As discussed in Chen and Chen (2008), if $\log p/\log n > 0.5$, the conventional BIC (Schwarz, 1978) is not consistent. In all the scenarios considered in this paper, $\log p/\log n > 1$. Our empirical studies confirm that in these scenarios the conventional BIC tends to be too liberal, and the extended BIC performs satisfactorily. The extended BIC for the linear model m is:

$$\text{BIC}_{\varrho}(m)=-2\log l_n\{\hat{\theta}(m)\}+df_m\log n+2\varrho\log\varsigma(S_{df_m}),$$

where df_m is the degrees of freedom for model m and $\varsigma(S_{df_m})$ is the number of the models containing df_m covariates. We take the number of the nonzero coefficient estimates in the model m as df_m and set $\varsigma(S_{df_m})=\begin{pmatrix} p \\ df_m \end{pmatrix}$, the number of combinations of df_m covariates chosen from p covariates. In addition, we set $\varrho \simeq 1 - 1/(2\log p/\log n)$ as $\varrho > 1 - 1/(2\log p/\log n)$ is suggested in Chen and Chen (2008). The extended BIC for a generalized linear model m is:

$$\text{BIC}_{\varrho}(m)=-2\log l_n\{\hat{\theta}(m)\}+df_m\log n+2df_m\varrho\log p,$$

where df_m is the number of nonzero coefficient estimates, and similar to the above $\rho \simeq 1 - 1/(2\log p/\log n)$, as suggested in Chen and Chen (2012).

4. Simulation

We evaluated those four penalties using a set of simulated data for multiple loci mapping problems. Specifically, the response variable is either a continuous trait (linear regression) or the case/control status (logistic regression), and the covariates are the genotypes of the SNPs. One particular challenge in a multiple loci mapping problem is that nearby SNPs often have correlated genotypes due to linkage disequilibrium, and such correlations may violate the irrepresentable condition, which is needed for the consistency of Lasso. To faithfully reproduce such correlation structure, we directly used genotype data of European Ancestry (EA) samples from a GWAS study of schizophrenia (Shi et al., 2009). The dataset was obtained from NCBI dbGaP, which includes GAIN (Genetic Association Information Network) samples (2,686/2,656: cases/controls, dbGaP Accession: phs000021.v3.p2) and non-GAIN samples (1,217/1,442: cases/controls, dbGaP Accession: phs000167.v1.p1) genotyped by Affymetrix 6.0 SNP arrays with ~900,000 SNPs.

To compare the performances of those penalty functions, we use two criteria to select the tuning parameters. One is the extended BIC as introduced earlier, and the other is an oracle criterion that uses the knowledge of the true model to select the tuning parameters. Certainly the oracle criterion is not applicable in practice when the true model is unknown. However, in simulation studies, the oracle criterion permits us to evaluate the performance of a penalty function rather than the combined outcome of a penalty function and a tuning parameter selection method. The oracle criterion is defined as follows. Let D be the number of discoveries, i.e., the covariates with non-zero regression coefficient estimates. D = TD + FD, where TD and FD are the number of true discoveries and false discoveries, respectively. The oracle criterion evaluates a model based on the three measures, the false discovery rate FD/D, power TD/s, and the sum of squared error of regression coefficient estimates $\sum_{j=1}^{p}|\hat{\beta}_j-\beta_{0j}|^2$, where β_{0j} is the true value of β_j. The model with the minimum of $\mathrm{wt}(\mathrm{FD/D-TD/s})+\sum_{j=1}^{p}|\hat{\beta}_j-\beta_{0j}|^2$ is selected, where wt is a weight to balance the number of true/false discoveries and bias. Models selected with larger wt tend to have more true discoveries and fewer false discoveries, but have a larger bias in their regression coefficient estimates.

4.1. Linear model

For computational efficiency when there are a large number of simulations, we randomly selected $n = 222$ samples and 12,656 SNPs with no missing values, and with a minor allele frequency greater than 5% on chromosome 20. The response variables y were simulated by $y = X\beta + \varepsilon$, where $\varepsilon \sim N(\mathbf{0}, I_{n\times n})$. We considered 3 situations involving different combinations of p and s: $p = 12{,}656$ and $s = 12$, 16, or 20. Let $u_1^\top = (0.5, -0.5, 0.4, -0.4)$. When $s = 12$, 16, and 20, β_0 are set by repeating u_1 three, four, and five times, respectively. In addition, we considered null situations with $s = 0$ and $p = 12{,}656$.

The tuning parameter grids were chosen as follows: a = (2.1, 2.5, 3.0, 3.7, 4.5, 6.0) for SCAD, a = (1.1, 2.0, 3.0, 4.0, 5.0, 6.0) for MCP, and 6 τ's for Log and SICA as described in the Supplementary Materials. We also applied Lasso implemented in R/glmnet. For each of these five penalties, 100 λ's uniformly distributed on a log scale were generated as described in the Supplementary Materials.

We used the extended BIC and oracle criteria $10(\mathrm{FD/D}-\mathrm{TD/s})+\sum_{j=1}^{p}|\hat{\beta}_j-\beta_{j0}|^2$ to select the tuning parameters. We give the term (FD/D – TD/s) a larger weight of 10 so that the oracle criterion selects the model with the smaller false discovery rate FD/D, greater power TD/s first, and use the sum of squared error of regression coefficient estimates $\sum_{j=1}^{p}|\hat{\beta}_j-\beta_{j0}|^2$ as a secondary criterion. Additional simulation results using various values of weight can be found in the Supplementary Materials.

For null simulation situations, all penalties have at most 1 or 2 false discoveries by the extended BIC tuning parameter selection criterion. Table 1 summarizes the simulation results in non-null situations with 12, 16, or 20 important covariates. The folded concave penalties perform better than the Lasso penalty. Among the four folded concave penalties, SICA, Log and MCP have comparable performance, and are better than SCAD when the tuning parameters are selected by the oracle criterion. When the tuning parameters are selected by the extended BIC, SICA and Log have comparable performance, and are better than SCAD and MCP. In additional simulation studies that are presented in the Supplementary Materials, SCAD and MCP with one tuning parameter (λ) have slightly worse performance than the situations with two tuning parameters. In contrast, Log and SICA with one tuning parameter (λ) have much worse performance than the situations with two tuning parameters. Therefore, the extra tuning parameter (a or τ) gives SCAD and MCP limited additional advantage, but significantly improves the performances of Log and SICA.

4.2. Simulation for logistic model

For penalized logistic regression, a larger sample size is needed for simulations with reasonable effect sizes. We randomly selected 10,156 SNPs (with a minor allele frequency larger than 5%) from chromosomes 1 to 22 and X and 750 samples (with a missing values percent smaller than 3%). We simulated the individual SNP effect so that the disease odds ratios are 2.0, corresponding to regression coefficients of 0.7. The binary response variables y were simulated based on the logistic regression model: $\log\{\Pr(y=1)/\Pr(y=0)\} = X\beta$, where s = 4, 8, or 12. In addition, the null model where s = 0 was simulated. The intercept was set as −2, corresponding to a disease prevalence of 0.12. The initial pool of tuning parameters were generated in the same way as linear regression, and then a particular combination of tuning parameters was selected to minimize the extended BIC, or an oracle criterion $10(\mathrm{FD/D}-\mathrm{TD}/s)+\sum_{j=1}^{p}|\hat{\beta}_j-\beta_{j0}|^2$.

For the simulation of null models, all penalties have at most 1 or 2 false discoveries by the extended BIC tuning parameter selection criterion. The simulation results of non-null models are shown in Table 2. In general, the results of logistic model simulation have a trend similar to that of linear model simulation. When the oracle criterion is used, all penalties

have satisfactory variable selection performances though SICA and Log have a smaller bias on effect size estimation. It can be observed that the models chosen by the oracle criterion are different from those selected by the extended BIC for SCAD and MCP. This is because the models chosen by the oracle criterion tend to have larger biases, which reduces the likelihood, and thus increases the realized value of the extended BIC. On the other hand, for Log and SICA, the models chosen by the oracle criterion are similar to those chosen by the extended BIC since they have a smaller bias on effect size estimation. Additional simulations presented in the Supplementary Materials confirm that SCAD with one or two tuning parameters have similar performance, and an additional tuning parameter improves MCP's performance. The additional tuning parameter significantly improves the performance of the SICA and Log penalties.

Finally, Table 3 presents the comparison of the computational burden for MCP, Log and SICA across various values of a and τ, respectively. It can be observed that the computation time of Log and SICA is much less than that of MCP.

In summary, Log and SICA have a smaller bias for the coefficient estimates of important covariates, and therefore, more accurate estimates of the likelihood function. In addition, they have lower computational burden compared to MCP. As a consequence, Log and SICA penalties have advantages in empirical usage.

5. Real data analysis

We analyzed the data of GWA studies of schizophrenia on European-ancestry samples (2,195 cases vs. 2,617 controls). The missing genotypic data were imputed using BEAGLE software (Browning and Browning, 2007), and 677,163 autosome SNPs with minor allele frequency no less than 5% were selected for the analysis. We included 23 principle components (PCs) of genotype data in the model to account for possible population stratification. First, a univariate logistic regression is conducted on the case-control status for each of the 677,163 SNPs, conditioning on the covariates: age, gender and 23 PCs. Using the resulting 677,163 p-values, we calculated a genomic control factor of 1.0445 (Devlin and Roeder, 1999), implying that there is no strong population stratification not accounted for in our model. The 7,984 SNPs with p-values smaller than 0.01 were selected for the following variable selection. We applied the penalized logistic regression on the 7,984 SNPs and 4,812 samples with the four folded-concave penalties, while accounting for the effects of age, gender and 23 PCs, by including them as unpenalized covariates.

We applied SCAD with $a = 3.7$ and MCP with $a = 3$, the default value of R package ncvreg, and chose to use two tuning parameters for SICA and Log. Using the extended BIC for tuning parameter selection, the penalized logistic regressions with Log and SICA selected 38 and 22 SNPs, respectively (Supplementary Table 1–2). However, penalized logistic regressions with both MCP and SCAD selected the null model since the null model has the lowest value of the extended BIC.

A joint model was fitted by a logistic regression using the 38 SNPs identified by the Log penalty together with age, gender, and 23 PCs to obtain the p-values for the 38 SNPs. The

results are illustrated in Figure 2, together with the marginal p-values for the 677,163 SNPs. There are 43 genes within 10kb distance of these 38 SNPs, and among them 21 are in the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang, Sherman and Lempicki, 2008). By functional category enrichment analysis at the DAVID website, 16 of the 21 genes are bound by transcription factor FOXO1, with significant enrichment p-value after a Benjamini correction. Recent studies have shown that FOXO1 regulates neuroblastoma differentiation (Mei et al., 2012), which is relevant to schizophrenia. In contrast, we also did the functional category analysis for those genes within 10 kb of the 38 SNPs with the smallest marginal p-values, but no functional category was significantly over-represented.

6. Conclusion and discussion

Although the methods with folded concave (nonconvex) penalties may not be desirable in terms of computational efficiency, they may lead to nice statistical properties in high dimensional setting (Fan and Li, 2001). To investigate the applicability of the nonconvex penalty functions in challenging high dimensional settings such as genomic studies, we conducted a theoretical analysis on the roles of tuning parameters with respect to the dimension of the problem and minimum effect size. The results suggest that the derivatives of the penalty function around 0 and the minimum effect size are two important quantities to be considered. A good performance of the penalized estimation requires that these two quantities be asymptotically different. Among the four penalties discussed in this paper, tuning one regularization parameter is sufficient to exploit the advantages of SCAD. In contrast, MCP, SICA and Log's performances can be significantly improved if two instead of one (λ) regularization parameter is tuned. These theoretical conclusions are well supported in the empirical studies. In the simulations, we also observe that a penalized estimation using SICA or Log appears to be computationally more efficient than using MCP. The good performance of tuning two regularization parameters comes with the cost of added computational time. In real data analysis, one needs to judge the difficulty of the penalization problem in terms of effect size and dimensionality in order to choose whether one or two regularization parameters are needed, and the theoretical results of this paper can guide such choices. These theoretical results are based on the sufficient conditions of the weak oracle property, and thus they could be refined if the sufficient and necessary conditions of the weak oracle property are available.

For the future work, it will be of great interest to study if the regularized methods using those four nonconvex penalties achieve feature selection consistency under the necessary and sufficient condition derived by Shen et al. (2013). Furthermore, Shen et al. (2013) have demonstrated that constrained approaches may offer both theoretical and computational advantages. Therefore, our following study may derive the constrained counterpart approaches for Log or SICA to enhance better empirical performances. In addition, Wang, Kim and Li (2013) have proposed an calibrated CCCP algorithm that produces a consistent solution path which contains the oracle estimator with probability approaching one. They also proposed a high-dimensional BIC criterion and showed that it can be applied to the solution path to select the optimal tuning parameter which asymptotically identifies the oracle estimator. Take the penalty SCAD at a fixed tuning value of $a = 3.7$ for example. The

calibrated CCCP algorithm introduces another parameter τ, and then two convex minimization problems using $\tau\lambda$ and λ are solved sequentially. For penalties that are sufficient to use one tuning parameter such as SCAD, the calibrated CCCP algorithm is ready to be applied. However, for the penalties required the usage of both of the tuning parameters such as Log penalty, it warrants future research on how to calibrate the two tuning parameters simultaneously in an efficient way. After the algorithm has been built, it will be interesting to see how different high-dimensional BIC criteria may influence the empirical performances and how the new devised methods perform in the genetic studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful for the constructive comments from the editors and the reviewers.

References

Breheny P, Huang J. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. The Annals of Applied Statistics. 2011; 5:232–253. [PubMed: 22081779]

Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. The American Journal of Human Genetics. 2007; 81:1084–1097. [PubMed: 17924348]

Bühlmann P, Mandozzi J. High-dimensional variable screening and bias in subsequent inference, with an empirical comparison. Computational Statistics. 2012:1–24.

Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. Biometrika. 2008; 95:759–771.

Chen J, Chen Z. Extended BIC for small-n-large-P sparse GLM. Statistica Sinica. 2012; 22:555.

Chen T-H, Sun W, Fine JP. Supplement to “Designing penalty functions in high dimensional problems: The role of tuning parameters”. 2016; doi: 10.1214/16-EJS1169SUPP

Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55:997–1004. [PubMed: 11315092]

Fan J. Comments on ‘Wavelets in statistics: A review’ by A. Antoniadis. Statistical Methods & Applications. 1997; 6:131–138.

Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association. 2001; 96:1348–1360.

Fan J, Lv J. A selective overview of variable selection in high dimensional feature space. Statistica Sinica. 2010; 20:101. [PubMed: 21572976]

Fan J, Lv J. Nonconcave Penalized Likelihood With NP-Dimensionality. Information Theory, IEEE Transactions on. 2011; 57:5467–5484.

Friedman, JH. Fast sparse regression and classification. 2008.

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols. 2008; 4:44–57.

Lv J, Fan Y. A unified approach to model selection and sparse recovery using regularized least squares. The Annals of Statistics. 2009; 37:3498–3528.

Mazumder R, Friedman JH, Hastie T. SparseNet: Coordinate descent with nonconvex penalties. Journal of the American Statistical Association. 2011; 106:MR2894769.

Mei Y, Wang Z, Zhang L, Zhang Y, Li X, Liu H, Ye J, You H. Regulation of neuroblastoma differentiation by fork-head transcription factors FOXO1/3/4 through the receptor tyrosine kinase PDGFRA. Proceedings of the National Academy of Sciences. 2012; 109:4898–4903.

Schwarz G. Estimating the dimension of a model. The Annals of Statistics. 1978; 6:461–464.

Shen X, Pan W, Zhu Y, Zhou H. On constrained and regularized high-dimensional regression. Annals of the Institute of Statistical Mathematics. 2013; 65:807–832. [PubMed: 24465052]

Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Peâ I, et al. Common variants on chromosome 6p22. 1 are associated with schizophrenia. Nature. 2009; 460:753–757. [PubMed: 19571809]

Sun W, Ibrahim JG, Zou F. Genomewide Multiple-Loci Mapping in Experimental Crosses by Iterative Adaptive Penalized Regression. Genetics. 2010; 185:349. [PubMed: 20157003]

Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological). 1996:267–288.

Wang L, Kim Y, Li R. Calibrating non-convex penalized regression in ultra-high dimension. Annals of statistics. 2013; 41:2505. [PubMed: 24948843]

Wright FA, Sullivan P, Brooks A, Zou F, Sun W, Xia K, Madar V, Abdellaoui A, Batista S, Butler C, Chen G, Chen TWC, et al. Heritability and Genomics of Gene Expression In Peripheral Blood. Nature Genetics. 2014 in press.

Zhang CH. Nearly unbiased variable selection under minimax concave penalty. The Annals of Statistics. 2010; 38:894–942.

Zhao P, Yu B. On model selection consistency of Lasso. The Journal of Machine Learning Research. 2006; 7:2541–2563.

Zou H, Li R. One-step sparse estimates in nonconcave penalized likelihood models. Annals of Statistics. 2008; 36:1509. [PubMed: 19823597]

Appendix

We present the following Theorem 1 of Fan and Lv (2011) for the self-completeness of this paper. This Theorem gives a set of sufficient and almost necessary conditions of a local maximizer of the penalized likelihood.

Theorem 1 (Characterization of PMLE). *$\hat{\beta} \in R^p$ is a strict local maximizer of the non-concave penalized likelihood $Q_n(\beta)=l_n(\beta)-\sum_{j=1}^{p} p_\varpi(|\beta_j|)$ if*

$$X_1^\top \mu(\hat{\theta})-X_1^\top y+np'_\varpi(\hat{\beta}_{01})=0 \quad (1)$$

$$\|X_2^\top (y-\mu(\hat{\theta}))\|_\infty-np'_\varpi(0+)<0 \quad (2)$$

$$\lambda_{\min}\left(X_1^\top \sum(\hat{\theta})X_1\right)-n\kappa(p_\varpi,\hat{\beta}_{01})>0. \quad (3)$$

The following Conditions 4.1–4.4 are for the design matrix X, and they are essentially the same as the corresponding conditions from Fan and Lv (2011). We first define a few notations used in the following regularity conditions. L_∞ norm of a matrix is the maximum of the L_1 norm of each row. $\lambda_{\max}()/\lambda_{\min}()$ denotes the maximum/minimum eigen-value of a

symmetric matrix, respectively. Denote a neighborhood of the non-zero coefficients as $\mathcal{N}_0 = \{\delta \in R^s : \|\delta - \beta_{01}\|_\infty \le d_n\}$.

Condition 4.1

$\|[X_1^\top \sum(\theta_0) X_1]^{-1}\|_\infty = O(b_s n^{-1})$, where

$$b_s = O(n^{\gamma_s}) \ll \min(n^{1/2-\gamma_0}, n^{\gamma_0-\nu}(\log n)^{-1/2}) \text{ and } \gamma_s \ge 0.$$

Condition 4.2

$\max_{\delta\in\mathcal{N}_0} \max_{j=1}^p \lambda_{\max}[X_1^\top |x_j| \mathrm{diag}\{|\mu''(X_1\delta)|\} X_1] = O(n)$, where the derivative $\mu''(X_1\delta)$ is taken component-wise.

Condition 4.3

$\max_{j=1}^p \|x_j\|_\infty = o(n^{(1-\alpha)/2}(\log n)^{-1/2})$ if the responses are unbounded.

Condition 4.4

$\max_{\delta\in\mathcal{N}_0} \kappa(p_\varpi, \delta) \le \min_{\delta\in\mathcal{N}_0} \lambda_{\min}[n^{-1} X_1^\top \sum(X_1\delta) X_1]$, where $\kappa(p_\varpi, \delta)$ is defined as the local concavity of a penalty function at $v = (v_1, \ldots, v_q)^\top$:

$$\kappa(p_\varpi, v) = \lim_{\varepsilon\to 0+} \max_{1\le j\le q} \sup_{t_1<t_2\in(|v_j|-\varepsilon, |v_j|+\varepsilon)} -\frac{p'_\varpi(t_2) - p'_\varpi(t_1)}{t_2 - t_1}.$$

For the penalties with continuous second derivatives,

$$\kappa(p_\varpi, v) = \max_{1\le j\le q} -p''_\varpi(v_j).$$

Given Conditions 1 to 4, we have the following weak oracle property.

Theorem 2 (Weak oracle property). *Given the Conditions 1 to 4, with probability at least* $P_{\text{converage}} = 1 - 2[sn^{-1} + (p-s)\exp(-n^a \log n)]$, *there exists a penalized likelihood estimator* $\hat\beta = (\hat\beta_1^\top, \hat\beta_2^\top)^\top$ *which satisfies* (*a*) *Sparsity*: $\hat\beta_2 = \mathbf{0}$, *(b)* L_∞ *loss*: $\|\hat\beta_1 - \beta_{10}\|_\infty = o(n^{-\gamma_0}\sqrt{\log n})$.

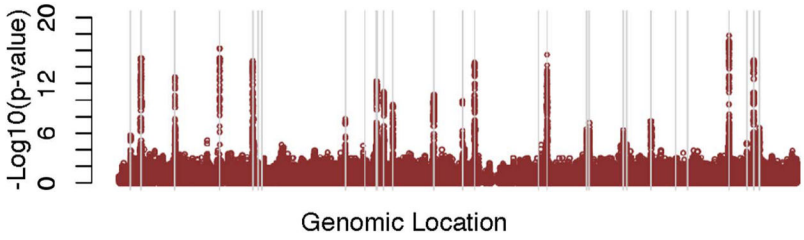

Fig 1.

Marginal association p-values for 645,316 SNPs on chromosome 1. The grey vertical lines denote the positions of 30 important SNPs. The genomic location spans 248,484,829 base-pairs. Note that a SNP is at a single base-pair location.

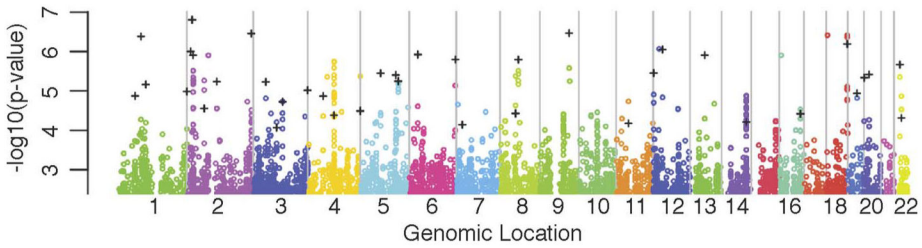

Fig 2.

GWA marginal p-values (colored circles) and the 38 SNPs (black crosses) identified by penalized logistic regression using Log penalty.

Table 1

Simulation results for penalized linear regression with (n=222, p = 12,656). The headers indicate the tuning parameter selection criterion (Oracle or the extended BIC) and the numbers in parentheses are the number of important covariates. For each penalty, we present the median of the number of true discoveries, false discoveries (in parentheses), and average bias of the true discoveries (in brackets) across 100 simulations.

	Oracle (12)	**Ext BIC (12)**	**Oracle (16)**	**Ext BIC (16)**	**Oracle (20)**	**Ext BIC (20)**
Lasso	11 (8) [0.33]	0 (0) [−]	7 (3) [0.39]	0 (0) [−]	14 (112) [0.34]	0 (0) [−]
SCAD	11 (3) [0.28]	0 (0) [−]	15 (25) [0.13]	0 (0) [−]	19 (27) [0.12]	0 (0) [−]
MCP	11 (1) [0.08]	10 (20) [0.08]	14 (2) [0.07]	11 (39) [0.10]	17 (3) [0.08]	5 (39) [0.11]
Log	11 (1) [0.07]	10 (3) [0.07]	14 (3) [0.07]	11 (7) [0.07]	17 (3) [0.08]	8 (10) [0.08]
SICA	11 (1) [0.06]	10 (3) [0.06]	14 (2) [0.06]	11 (6) [0.07]	17 (4) [0.07]	5 (7) [0.08]

Table 2

Simulation results for penalized logistic regression (n=750, p = 10,156). The headers indicate the tuning parameter selection criterion (Oracle or the extended BIC) and the numbers in parentheses are the number of important covariates. For each penalty, we present the median of the number of true discoveries, the number of false discoveries (in parentheses), and the average bias of true discoveries (in brackets) across 100 simulations.

	Oracle (4)	**Ext BIC (4)**	**Oracle (8)**	**Ext BIC (8)**	**Oracle (12)**	**Ext BIC (12)**
Lasso	4(0) [0.49]	4 (0) [0.47]	7(0) [0.55]	6 (0) [0.53]	11(2) [0.59]	0 (0) [−]
SCAD	4 (0) [0.48]	4 (0) [0.39]	7 (0) [0.53]	6 (0) [0.43]	11(2) [0.58]	0 (0) [−]
MCP	4 (0) [0.093]	4 (0) [0.097]	7 (0) [0.25]	6 (1) [0.14]	11(1) [0.32]	11 (7) [0.25]
Log	4 (0) [0.085]	4 (0) [0.096]	7 (0) [0.085]	7 (1) [0.09]	11(1) [0.10]	11 (1) [0.10]
SICA	4 (0) [0.084]	4 (0) [0.094]	7 (0) [0.095]	7 (1) [0.099]	11(1) [0.12]	11 (1) [0.096]

Table 3

Running time rounded to minutes per simulation (n=750, s = 12, p = 10,156) for 100 λ's and a fixed a of MCP or τ of Log and SICA.

MCP	21.1 ($a = 1.1$)	5.2 ($a = 2.0$)	7.1 ($a = 3.0$)	6.3 ($a = 4.0$)	9.7 ($a = 5.0$)
Log	2.1 ($\tau = 10^{-6}$)	1.9 ($\tau = 10^{-5}$)	1.9 ($\tau = 10^{-4}$)	1.9 ($\tau = 10^{-3}$)	1.8 ($\tau = 0.6$)
SICA	2.0 ($\tau = 10^{-6}$)	2.1 ($\tau = 10^{-5}$)	1.9 ($\tau = 10^{-4}$)	1.8 ($\tau = 10^{-3}$)	1.8 ($\tau = 0.6$)

Electron J Stat. Author manuscript; available in PMC 2017 October 05.