IDENTIFICATION OF SIGNIFICANT GENETIC VARIANTS VIA *SLOPE*, AND ITS EXTENSION TO *GROUP SLOPE*

ALEXEJ GOSSMANN[1], SHAOLONG CAO[2], YU-PING WANG[2,3]

ACM CONFERENCE ON BIOINFORMATICS, COMPUTATIONAL BIOLOGY, AND HEALTH INFORMATICS

2015/9/10

---

1. Department of Mathematics, Tulane University, New Orleans

2. Department of Biomedical Engineering, Tulane University, New Orleans

3. Department of Biostatistics and Bioinformatics, Tulane University, New Orleans

INTRODUCTION

THE MODEL SELECTION PROBLEM

- Linear model $\mathbf{y} = X\mathbf{b} + \mathbf{z}$, where $\mathbf{y} \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, $\mathbf{b} \in \mathbb{R}^p$, $\mathbf{z} \sim \mathrm{N}(0, \sigma^2 I)$.
- Possibly $n < p$.
- *Estimation*: Find best predictions for $\mathbf{y}$ or $\mathbf{b}$.
- *Feature selection*: Find which b_i are non-zero.

INTRODUCTION

THE MODEL SELECTION PROBLEM IN GENETICS

- Genomic, proteomic, epigenomic, metabolomic, etc. data are typically high-dimensional and suffer from the curse of dimensionality.
- Elimination of noisy or redundant features leads to more accurate prediction.
- Prediction of a disease phenotype based on a handful of features is needed for inexpensive diagnosis.
- Feature selection can lead to better understanding of the underlying biology.

ℓ_0 REGULARIZATION (E.G. C_p BY MALLOWS, 1973, AND AIC BY AKAIKE, 1974)

$$\min_{\mathbf{b}\in\mathbb{R}^p} \|\mathbf{y} - X\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_0$$

- ℓ_0 norm is non-convex $\rightsquigarrow$ Not practical for large p (e.g. for $p = 100$)

ℓ_1 REGULARIZATION (E.G. LASSO BY TIBSHIRANI, 1994)

$$\min_{\mathbf{b}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_1$$

- Small λ leads to the selection of too many irrelevant parameters (ineffective in sparse settings).
- Large λ yields little power as well as a large bias.

SLOPE

SORTED L-ONE PENALIZED ESTIMATION (BOGDAN, VAN DEN BERG, SU, CANDES, 2013)

$$\min_{\mathbf{b}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\mathbf{b}\|_2^2 + \sum_{i=1}^{p} \lambda_i |\mathbf{b}|_{(i)}$$

- *Regularizing sequence* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$
- $|b|_{(1)} \geq |b|_{(2)} \geq \ldots \geq |b|_{(p)}$ denotes the order statistic of the magnitudes of the vector $\mathbf{b} \in \mathbb{R}^p$

- M. Bogdan, E. van den Berg, W. Su, and E. Candes. *Statistical estimation and testing via the sorted L1 norm*. ArXiv e-prints, Oct. 2013.
- M. Bogdan, E. van den Berg, C. Sabatti, W. Su, and E. J. Candes. *SLOPE – Adaptive Variable Selection via Convex Optimization*. ArXiv e-prints, July 2014.
- E. Candes and W. Su. *SLOPE is Adaptive to Unknown Sparsity and Asymptotically Minimax*. ArXiv e-prints, Mar. 2015.

SLOPE

- SLOPE is convex.
- Computational cost is roughly the same as for the LASSO.
- Adaptivity to the sparsity level: the cost of including new variables decreases as more variables are added to the model.
- Related to the BHq procedure (Y. Benjamini and Y. Hochberg, 1995) with similar FDR control properties.

FALSE DISCOVERY RATE (FDR)

Essentially, the SLOPE procedure is testing the p hypotheses $H_i : b_i = 0$ for $i = 1, \ldots, p$, where H_i is rejected iff $\hat{b}_i \neq 0$.

SLOPE aims to control the FDR, i.e. the proportion of the irrelevant among all selected predictors.

ORTHOGONAL DESIGNS

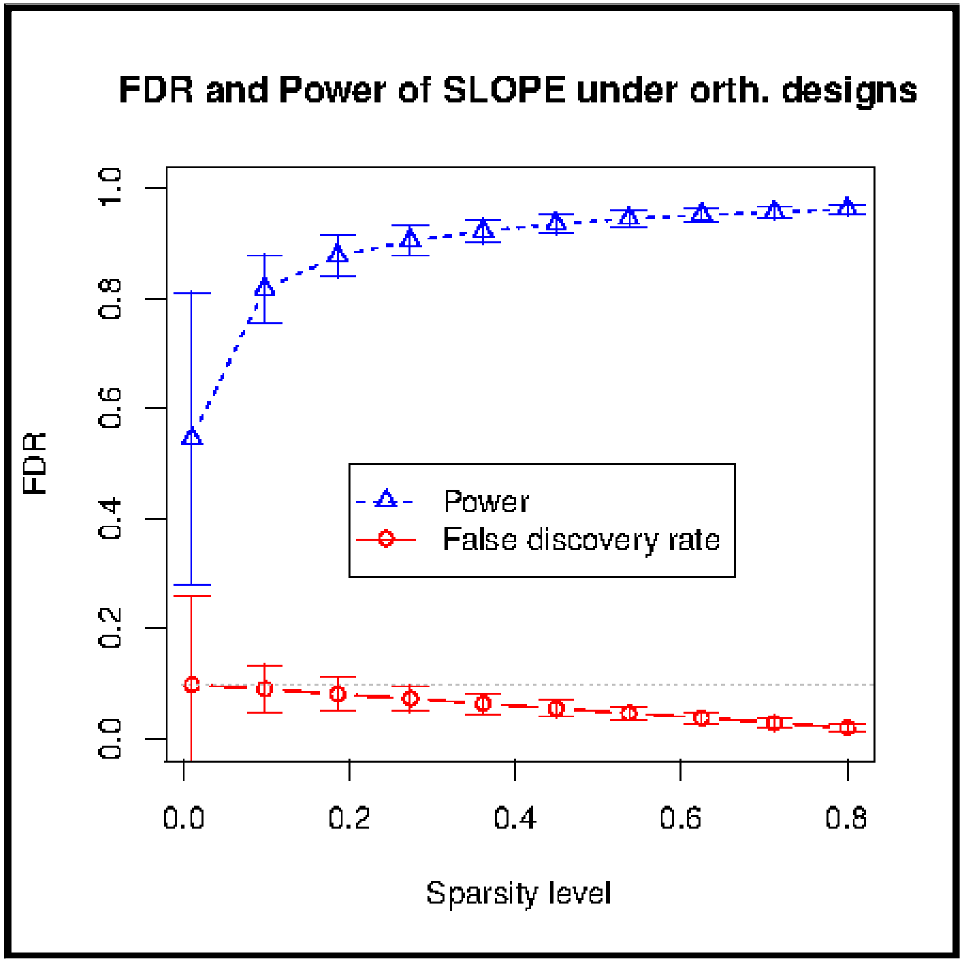
Let R = #rejections, V = #false rejections, p_0 = #true null hypotheses.

THEOREM (BOGDAN, VAN DEN BERG, SU, CANDES, 2013)

Assume an orthogonal design with i.i.d. $\mathrm{N}(0, 1)$ errors, and set $\lambda_i = \Phi^{-1}\left(1 - q\frac{i}{2p}\right)$. Then the FDR of SLOPE obeys

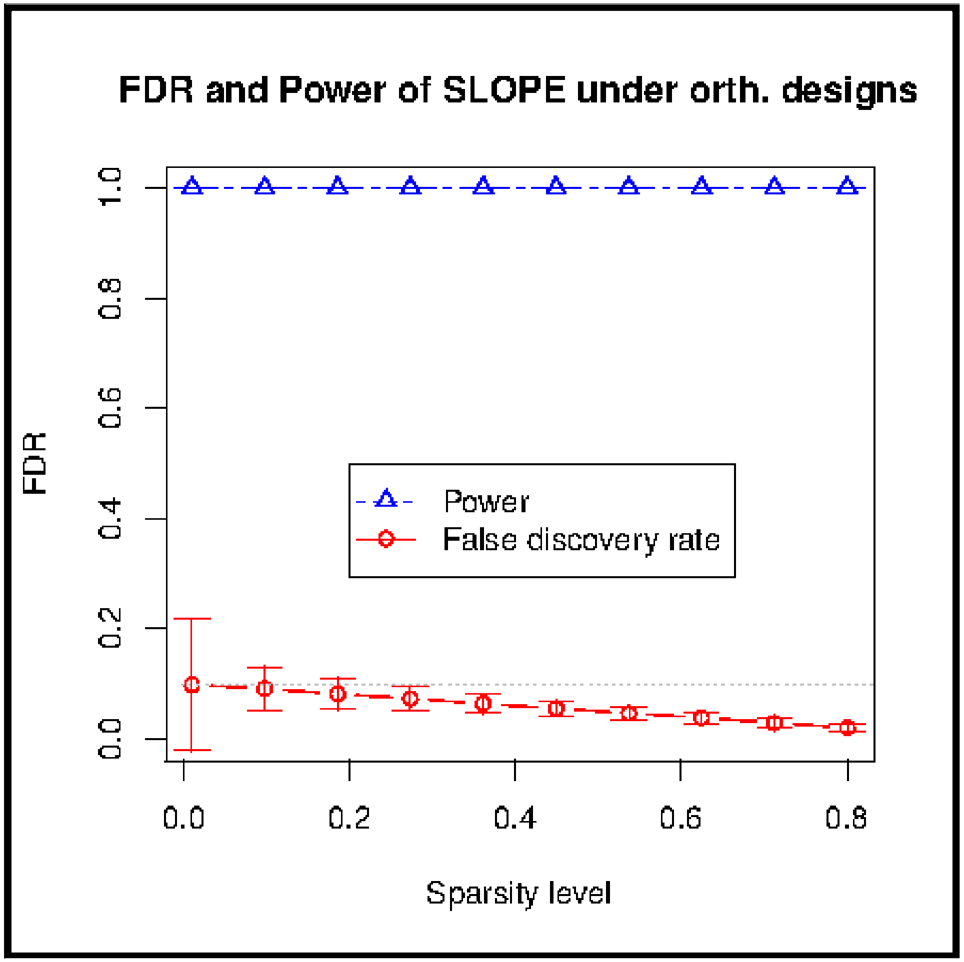
$$\mathrm{FDR} = \mathrm{E}\left(\frac{V}{\max(R, 1)}\right) \leq q\frac{p_0}{p}.$$

ORTHOGONAL DESIGNS



500×500 orthogonal design, $\sqrt{2\log(n)}$ signal strength, 1000 replications at each sparsity level, bars for $\pm$ SD

ORTHOGONAL DESIGNS



$5\sqrt{2\log(n)}$ signal strength (5 times previous)

NONORTHOGONAL DESIGNS

- Theorem is not valid for nonorthogonal design matrices.
- The regularizing sequence can be adjusted:

$$\begin{aligned} \lambda_1 &= \lambda_1^{(\mathrm{BH})}, \\ \lambda_i &= \lambda_i^{(\mathrm{BH})} \sqrt{1 + \omega(i-1)}, \end{aligned}$$

where $\lambda_i^{(\mathrm{BH})} = \Phi^{-1}\left(1 - q\frac{i}{2p}\right)$ and

$\omega(i) \approx \mathrm{E}\left[\left(X_i^T X_S (X_S^T X_S)^{-1} \lambda_S\right)^2\right]$ with $S = \mathrm{supp}(\mathbf{b})$.

- $\omega(i)$ can be approximated with a Monte Carlo simulation.

UNKNOWN NOISE LEVEL AND INTERCEPT

- SLOPE does not include an intercept term and presupposes the knowledge of the noise level σ^2.
- Estimation of the intercept can be avoided by standardizing the response as well as the predictor variables.
- σ^2 can be estimated by the following iterative procedure:

1. Set $\hat{\sigma}^{(0)}$ equal to the sample standard deviation of $\mathbf{y}$.
2. Update $\hat{\sigma}^{(k)}$ using linear regression on $\mathrm{supp}\left(\hat{\mathbf{b}}^{(k-1)}\right)$, which is identified by SLOPE with $\hat{\sigma}^{(k-1)}$.
3. Repeat step 2 until $\mathrm{supp}\left(\hat{\mathbf{b}}^{(k)}\right) = \mathrm{supp}\left(\hat{\mathbf{b}}^{(k-1)}\right)$.

APPLICATION TO GENETICS

SIMULATION OF REALISTIC DNA SEQUENCE DATA

- SeqSIMLA2 (Chung et al. 2015) and cosi (Schaffner et al. 2005) were used to simulate DNA sequence data that closely resemble empirical data.
- Each of 100 simulated data sets consists of 5330 SNPs (single nucleotide polymorphism) for 2000 unrelated individuals.
- The phenotype is a quantitative trait simulated under the additive model in SeqSIMLA2.
- Significant SNPs were randomly selected among SNPs with MAF (minor allele frequency) of at least 0.01

APPLICATION TO GENETICS

SIMULATION OF REALISTIC DNA SEQUENCE DATA

We consider two scenarios:

1. 5 significant SNPs, each explaining 10% of the phenotypic variance; the remaining 50% of the variance due to environmental effects; no polygenic effects.
2. 20 significant SNPs, each explaining 5% of the variance; no environmental or polygenic effects.

APPLICATION TO GENETICS

DATA PRUNING

- Even with $\lambda_1, \lambda_2, \ldots, \lambda_p$ adjusted as described previously, SLOPE cannot handle high correlations between predictors well.
- Data is pruned such that the maximal pair-wise correlation between predictors does not exceed 0.3, by iteratively removing columns from the design matrix based on their average pair-wise correlation and their univariate association with the response.

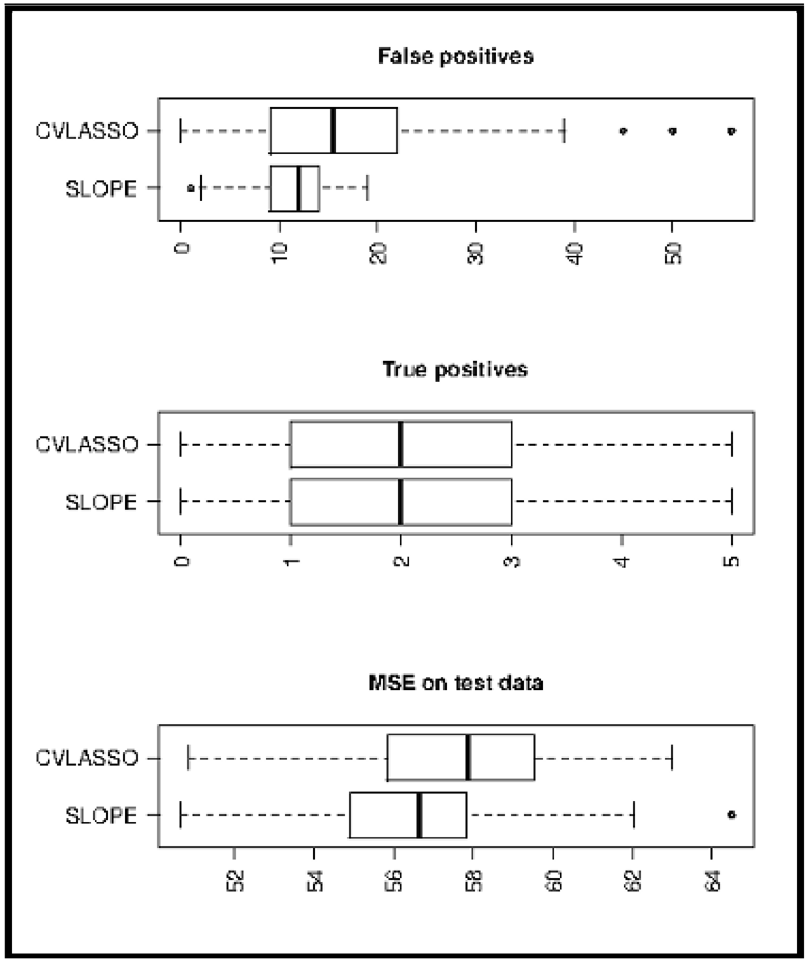
$\rightsquigarrow$ Of the 5330 SNPs approximately 320 remain in the data, and approximately half of the significant SNPs are discarded...

APPLICATION TO GENETICS

RESULTS

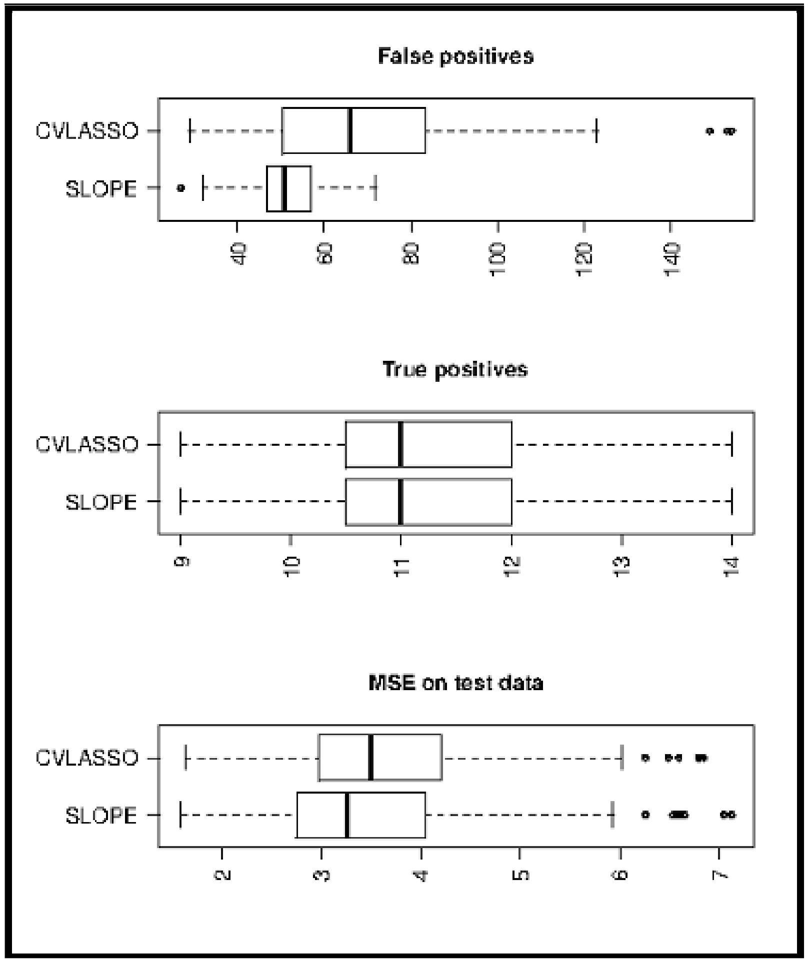
- We compare the performance of SLOPE to the LASSO.
- The LASSO regularization parameter λ is selected by ten-fold cross-validation.

APPLICATION TO GENETICS RESULTS



5 significant SNPs, each explaining 10% of the phenotypic variance

APPLICATION TO GENETICS RESULTS



20 significant SNPs, each explaining 5% of the phenotypic variance

GROUP SLOPE MOTIVATION

- SLOPE works best if the predictor variables have very small pair-wise correlations.
- Typically, genetic data is highly correlated.

$\Rightarrow$ Genetic data needs to be pruned to a great extent, in order to get good results with SLOPE.

GROUP SLOPE MOTIVATION

- Often the data can be subdivided into groups with possibly a high within group correlation but a low between group correlation.
- Specifically in genomic data analysis, SNPs in a gene or genes in a pathway can be available as prior knowledge along with a sparsity assumption.

$\Rightarrow$ Select or drop entire groups rather than individual significant predictors.

GROUP LASSO (M. YUAN AND Y. LIN, 2006, AND OTHERS)

- $\mathbf{y} = X\mathbf{b} + \mathbf{e}, \mathbf{y} \in \mathbb{R}^n, X \in \mathbb{R}^{n \times p}, \mathbf{b} \in \mathbb{R}^p, \mathbf{e} \sim \mathrm{N}(0, \sigma_e^2 I)$.
- The predictor variables $\mathbf{b}$ are divided into J groups of sizes $p_1, p_2, \cdots, p_J$, i.e. $\mathbf{b} = (\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots, \mathbf{b}_J^T)^T$ with $\mathbf{b}_i \in \mathbb{R}^{p_i}$.
- Estimate $\mathbf{b}$ as the solution to the convex minimization problem

$$\min_{\mathbf{b} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{y} - X\mathbf{b}\|_2^2 + \sum_{i=1}^{J} \lambda_i \sqrt{p_i} \|\mathbf{b}_i\|_2 .$$

- For any i this procedure either keeps the entire block $\mathbf{b}_i$ non-zero, or sets all its components to zero.

GROUP SLOPE MODEL

- Group SLOPE is related to Group LASSO in the same way in which SLOPE is related to LASSO.
- Define the Group SLOPE minimization problem as

$$\min_{\mathbf{b}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - X\mathbf{b}\|_2^2 + \sum_{i=1}^{J} \lambda_i \sqrt{p_{(i)}}\,\|\mathbf{b}_{(i)}\|_2,$$

where $\sqrt{p_{(1)}}\,\|\mathbf{b}_{(1)}\|_2 \geq \sqrt{p_{(2)}}\,\|\mathbf{b}_{(2)}\|_2 \geq \ldots \geq \sqrt{p_{(J)}}\,\|\mathbf{b}_{(J)}\|_2.$

COMPUTATIONAL ALGORITHMS

- A group-wise generalization of the algorithm in the original SLOPE paper (Bogdan, van den Berg, Su, Candes, 2013).
- The minimization problem can be rewritten as a sum of a convex function and a differentiable convex function with a Lipschitz continuous derivative:

$$\min_{\mathbf{c}\in\mathbb{R}^p} f_1(\mathbf{c}) + f_2(\mathbf{c}),$$

$$\begin{aligned}
f_1(\mathbf{c}) &= \frac{1}{2}\left\|\mathbf{y} - XD^{-1}\mathbf{c}\right\|_2^2, \\
f_2(\mathbf{c}) &= \sum_{i=1}^{J} \lambda_i \left\|\mathbf{c}_{(i)}\right\|_2, \\
\mathbf{c}_i &= \sqrt{p_i}\,\mathbf{b}_i.
\end{aligned}$$

PROXIMAL GRADIENT METHOD FOR GROUP SLOPE

$\varepsilon \in \left(0, \min\left(1, \frac{1}{\xi}\right)\right), \mathbf{b}^{(0)} \in \mathbb{R}^p, \mathbf{c}^{(0)} = D\mathbf{b}^{(0)}$

for $k = 0, 1, 2, \dots$ do

$$\begin{aligned} \gamma_k &\in \left[\varepsilon, \frac{2}{\xi} - \varepsilon\right] \\ \mathbf{c}^{(k+1)} &\leftarrow \operatorname{prox}_{\gamma_k f_2}\left(\mathbf{c}^{(k)} - \gamma_k \left(XD^{-1}\right)^T \left(X\mathbf{b}^{(k)} - \mathbf{y}\right)\right) \\ \mathbf{b}^{(k+1)} &= D^{-1}\mathbf{c}^{(k+1)} \end{aligned}$$

end for

COMPUTING THE PROX

Proximal mapping:

$$\operatorname{prox}_{f_2}(y) = \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^p} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \sum_{i=1}^{J} \lambda_i \left\|\mathbf{x}_{(i)}\right\|_2 .$$

COMPUTING THE PROX

LEMMA

If $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_J)^T \in \mathbb{R}^J$ is the solution of the minimization problem

$$\min_{\tilde{\mathbf{x}} \in \mathbb{R}^J} \frac{1}{2} \sum_{i=1}^{J} \left(\|\mathbf{y}_i\|_2 - \tilde{x}_i \right)^2 + \sum_{i=1}^{J} \lambda_i |\tilde{x}|_{(i)}.$$

Then the solution to $\mathrm{prox}_{f_2}(y)$ is given by

$$\begin{aligned} \mathbf{x} &= (\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_J^T)^T \text{with} \\ \mathbf{x}_i &= \frac{\tilde{x}_i}{\|\mathbf{y}_i\|_2} \mathbf{y}_i, \quad \forall i \in 1, \ldots, J, \end{aligned}$$

where $\mathbf{y}_i \in \mathbb{R}^{p_i}$ denotes the ith block of $\mathbf{y} \in \mathbb{R}^p$ for $i \in 1, \ldots, J$.

COMPUTING THE PROX

The Lemma combined with the fast prox algorithm for the SLOPE method (Algorithm 4 in Bogdan, van den Berg, Sabatti, Su, Candes, 2014) implies a simple algorithm for the prox function.

ALGORITHM COMPUTING THE PROX

$\mathbf{x} = (\mathbf{x}_1^T, \ldots, \mathbf{x}_J^T)^T$

$\mathbf{y} = (\mathbf{y}_1^T, \ldots, \mathbf{y}_J^T)^T$

$\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_J)^T = (\|\mathbf{y}_1\|_2, \|\mathbf{y}_2\|_2, \ldots, \|\mathbf{y}_J\|_2)^T$

$\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_J)^T = \mathrm{prox}_{J_\lambda}(\tilde{\mathbf{y}})$

`for` $k = 1, 2, \ldots, J$ `do`

$\quad \mathbf{x}_i = \frac{\tilde{x}_i}{\tilde{y}_i}\mathbf{y}_i$

`end for`

where $\mathrm{prox}_{J_\lambda}$ is the prox function of SLOPE.

REGULARIZING SEQUENCE

- In order to (approximately) control the false discovery rate, we need to select suitable $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_J$.
- Can procedures available for the SLOPE method be generalized for Group SLOPE?

REGULARIZING SEQUENCE

A SIMPLIFIED SPECIAL CASE

- Assume that the columns of X are all equal within each block but different between different blocks.
- Collapse $X \in \mathbb{R}^{n\times p}$ into $\tilde{X} \in \mathbb{R}^{n\times J}$, and let $\tilde{\mathbf{b}} \in \mathbb{R}^J$ have entries $\tilde{b}_i = p_i \mathbf{b}_{i\,1}$. Then the objective function becomes:

$$\frac{1}{2}\left\|\mathbf{y} - \tilde{X}\tilde{\mathbf{b}}\right\|_2^2 + \sum_{i=1}^{J} \lambda_i \left|\tilde{b}_{(i)}\right|.$$

- This has the form of the regular SLOPE problem, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_J$ can be constructed by the available procedure.

REGULARIZING SEQUENCE

For a general model matrix X the above motivates the following approach:

1. Construct a matrix $\tilde{X}$ by taking its ith column to be the average of the columns of the ith block of X.
2. Normalize the columns of $\tilde{X}$ to have norms equal to one.
3. Construct a regularizing sequence $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_J$ using the Monte Carlo based method for SLOPE.

SIMULATION RESULTS

SIMULATED DATA

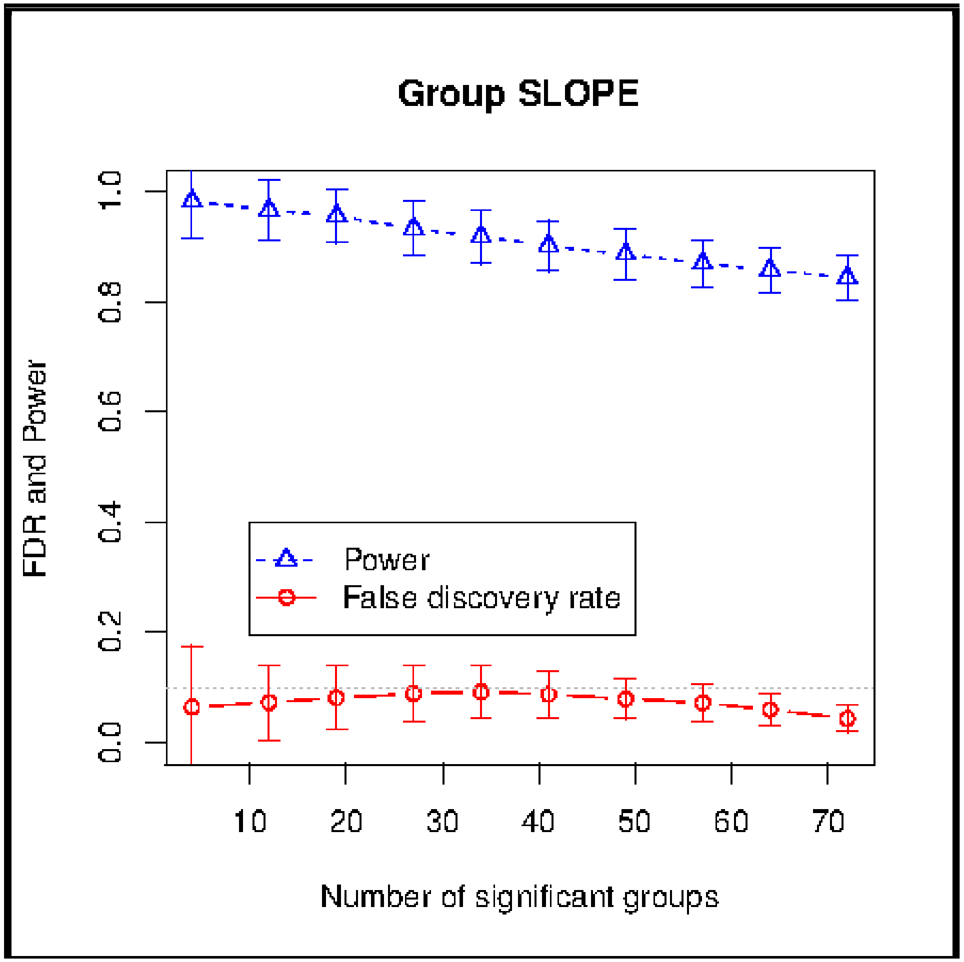
- $n = 200, p = 1050, \mathbf{y} = X\mathbf{b} + \mathbf{e}$ with $\mathbf{e} \sim \mathrm{N}(0, I)$.
- The p predictors are divided into 90 groups; 30 groups of size 5, 30 groups of size 10, and 30 groups of size 20.
- The non-zero variables are set to be ± 1 (same sign within a block).

SIMULATION RESULTS

SIMULATED DATA

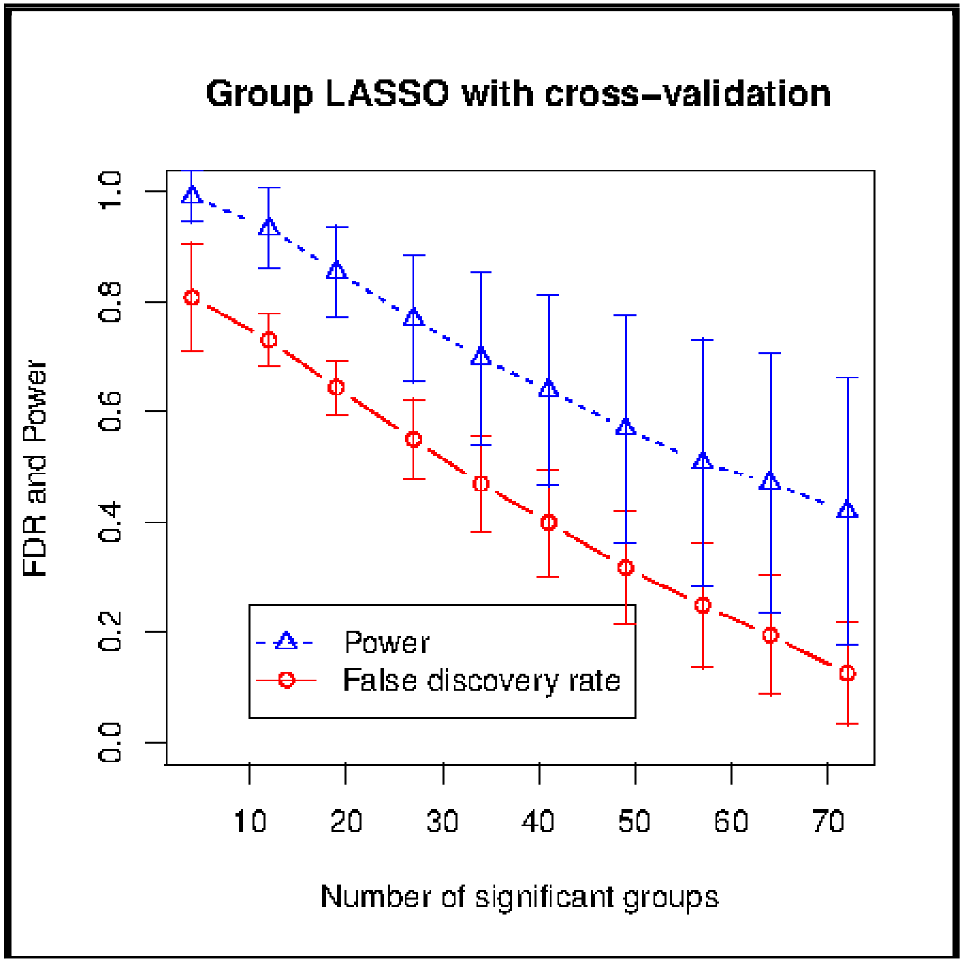
- We consider ten sparsity levels (proportion of significant groups among their total number).
- At each sparsity level we consider:
  - A case with very high within group correlations (≈ 0.99) and very low between group correlations (≈ 0.05).
  - A setting with only moderately large within group correlations (≈ 0.7) and moderate between group correlations (≈ 0.3).
- At each sparsity level 1000 repetitions are performed for each setting.

SIMULATION RESULTS



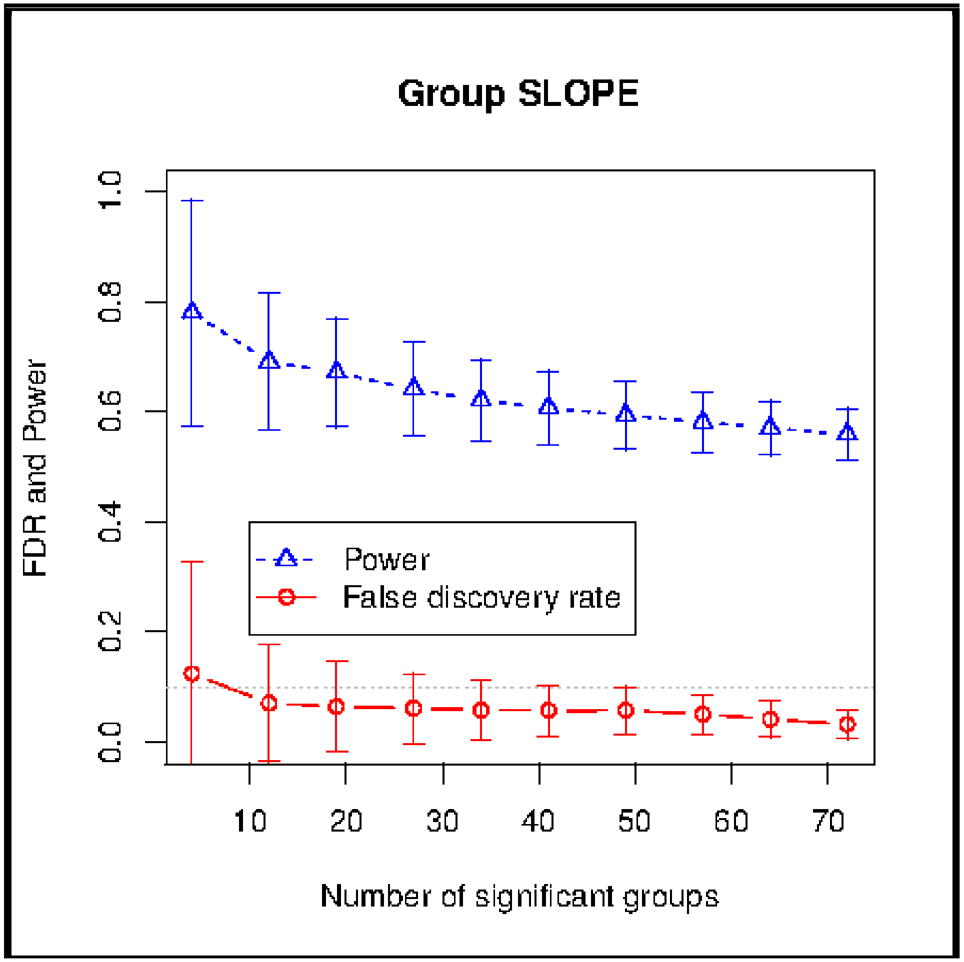
- within group correlations ≈ 0.99
- between group correlations ≈ 0.05
- bars correspond to $\pm$ SD

SIMULATION RESULTS



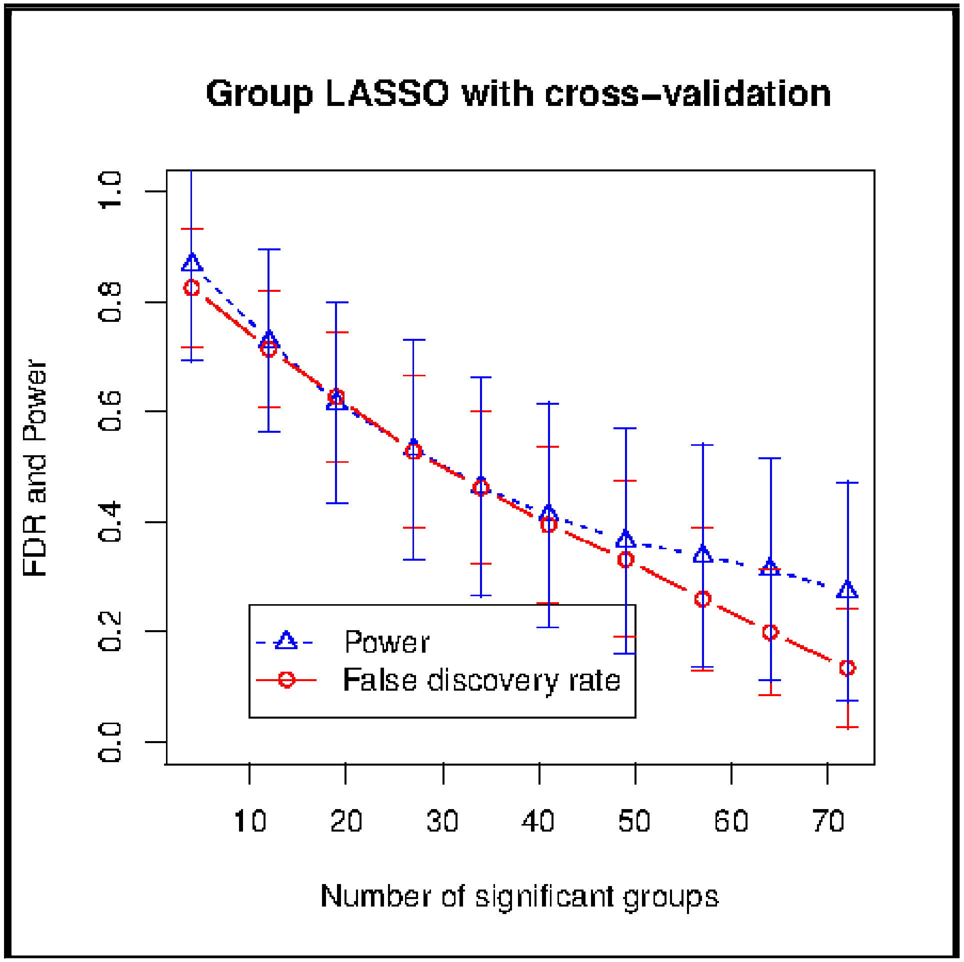
- within group correlations ≈ 0.99
- between group correlations ≈ 0.05
- bars correspond to $\pm$ SD

SIMULATION RESULTS



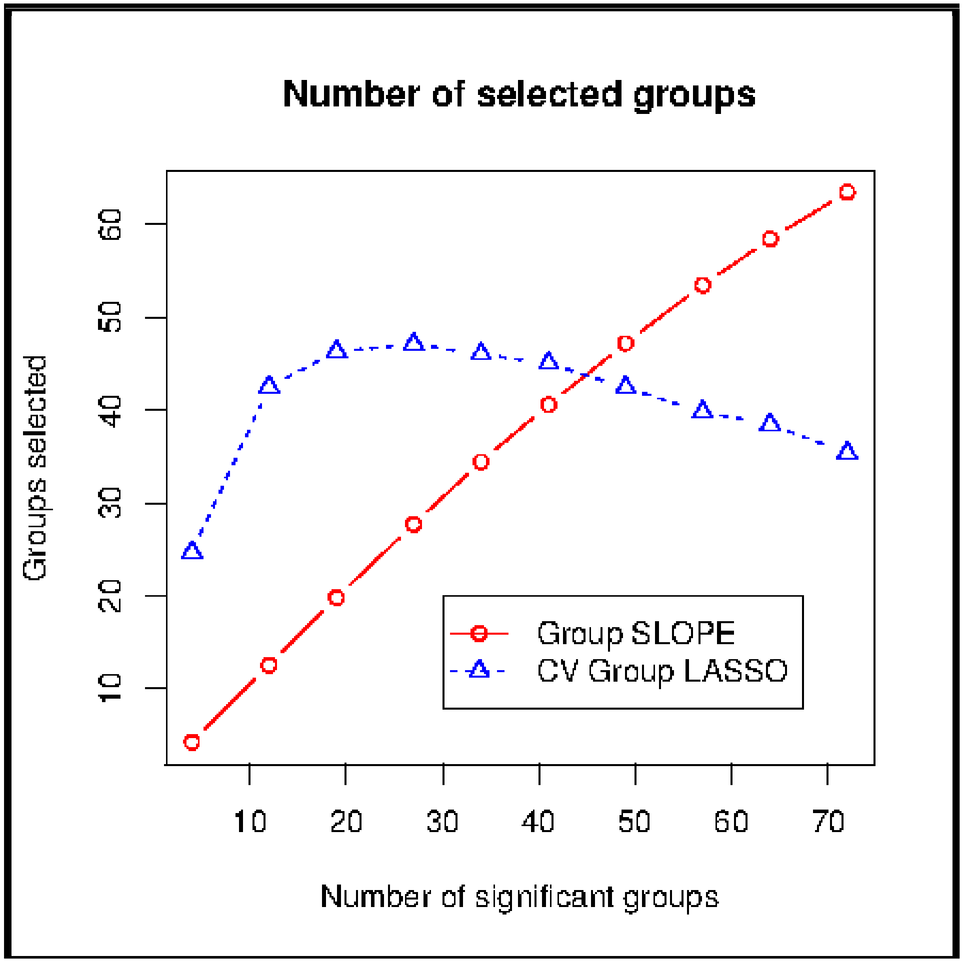
- within group correlations ≈ 0.7
- between group correlations ≈ 0.3
- bars correspond to $\pm$ SD

SIMULATION RESULTS



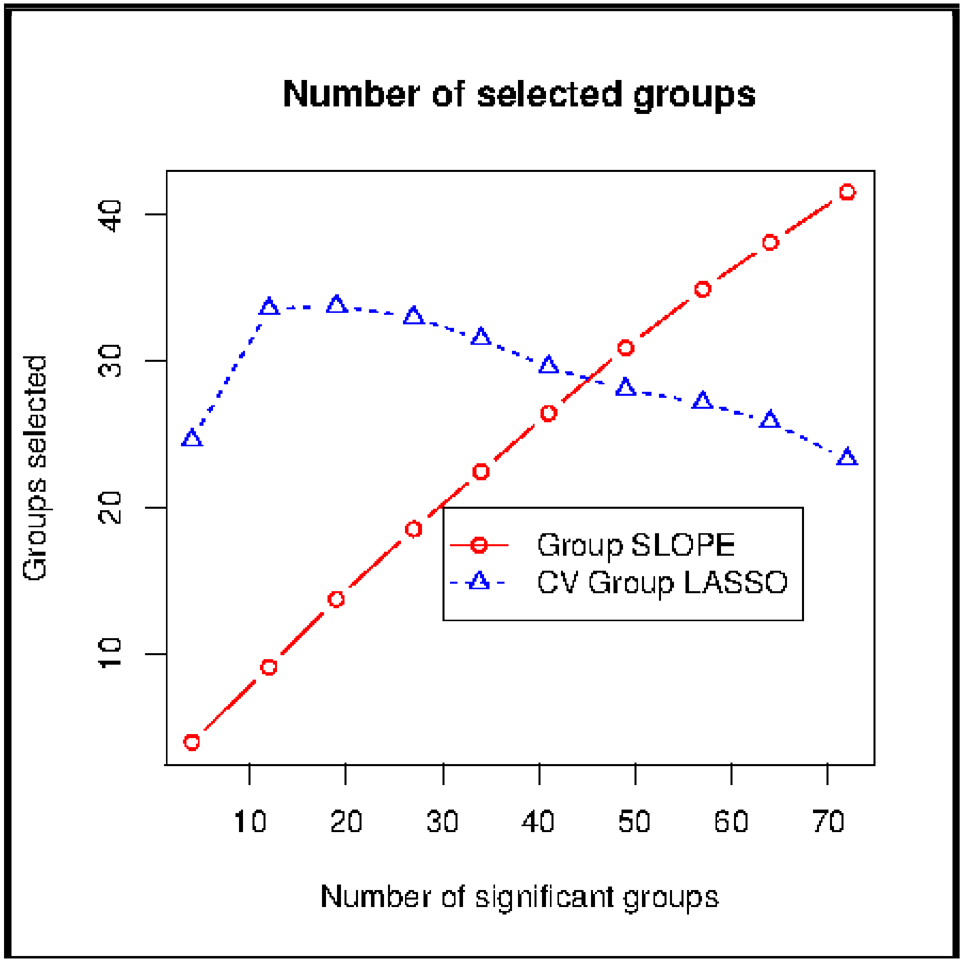
- within group correlations ≈ 0.7
- between group correlations ≈ 0.3
- bars correspond to $\pm$ SD

SIMULATION RESULTS



- within group correlations ≈ 0.99
- between group correlations ≈ 0.05

SIMULATION RESULTS



- within group correlations ≈ 0.7
- between group correlations ≈ 0.3

CONCLUSION

- SLOPE outperformed LASSO in terms of FDR as well as prediction MSE while having the same detection power.
- However, FDR of the SLOPE largely exceeded the nominal level of 0.1. Possibly data simulated by SeqSIMLA does not match SLOPE in some way.

CONCLUSION

- For very sparse data (sparsity level < 0.1), even under orthogonal designs the false discovery proportion is quite unstable, and often exceeds the aimed level significantly in our simulations.
- Same appears to be true for Group SLOPE.
- In many genomic instances the solution resides at these very sparse levels. This might require special care in future applications.

CONCLUSION

Similar to SLOPE, in considered settings...

- Group SLOPE adapts the number of selected groups to the unknown true number of significant groups of predictors.
- Group SLOPE keeps the false discovery rate below a specified level.
- Group LASSO has a much higher FDR and a lower detection power than Group SLOPE.

FUTURE WORK

- Application to real data
- Effect of covariates of different directionality in the same block in the Group SLOPE model
- Different ways of dividing the data into blocks
- Incorporation of other types of prior knowledge, e.g. family relationships among the subjects as random effects in the model
- Ways to construct the regularizing sequence $\lambda_1, \lambda_2, \ldots, \lambda_p$, which are less computationally expensive than the Monte Carlo approach

ACKNOWLEDGMENTS

- Our work is partially supported by NIH R01 GM109068 and R01 MH104680.
- We would also like to thank Malgorzata Bogdan* and Weijie Su** (two of the original authors of SLOPE) for helpful comments.

---

* Wroclaw University of Technology, Department of Mathematics and Computer Science
** Stanford University, Department of Statistics