SLOPE-Adaptive variable selection via convex optimization

Guidelines in ML

Nafissa BENALI & Angel REYERO

M2 Mathematics & Artificial Intelligence
Institut de Mathématiques d'Orsay (IMO)
Paris-Saclay University

Teachers:
Claire BOYER, Olivier COUDRAY, Guillermo DURAND

28 March, 2024

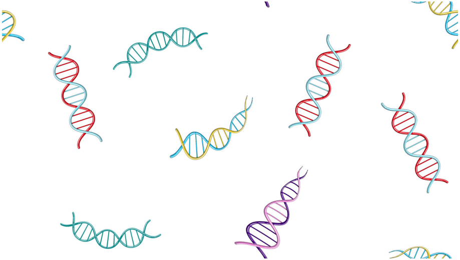
Contents

Index

Motivations

Guarentees on variable selection

We aim to select variables based on their importance relative to the target variable, rather than solely for optimal prediction.



A common starting point for exploring relationships between covariates X and a target variable Y is the linear model:

$$\underbrace{Y}_{\text{cholesterol}} = \underbrace{X}_{\text{genes}} \beta + Z$$

where $Z \sim \mathcal{N}(0, \sigma^2 I_n)$ and $X \in \mathscr{M}_{n,p}(\mathbb{R})$ In our example we have p $\gg$ n.

Quick reminder

Common metrics

The challenge is to consider metrics and methods that enable us to achieve our goal of ensuring finite sample guarentees for the selected subsets of variables. Let $\mathscr{R}$ denote the set of rejected variables, and let $V(\mathscr{R}) = |\mathscr{R} \cap \mathscr{H}_0|$ represent the cardinality of the false positives set.

Common metrics	Methods
FWER, $\mathbb{P}(\lvert\mathscr{R} \cap \mathscr{H}_0\rvert \geq 1)$	Bonferroni, $\mathscr{R} = \{i; p_i \leq \frac{\alpha}{p}\}$
FDR, $\mathbb{E}\left[\frac{V(\mathscr{R})}{\lvert\mathscr{R}\rvert \vee 1}\right]$	Benjamini-Hochberg, $\mathscr{R} = \{i : p_i \leq \frac{\alpha \widehat{k}^{\mathrm{BH}}}{p}\}$, where $\widehat{k}^{\mathrm{BH}} = \max\left\{k \in [\![0, p]\!]; p_{(k)} \leq \frac{\alpha k}{p}\right\}$

Index

Methodology

Introduction: Lasso is commonly used for variable selection. In our study, we will explore variations of Lasso.

Lasso:

$$\min_{b\in\mathbb{R}^p} \frac{1}{2}\|Y - Xb\|_2^2 + \lambda\|b\|_1 \tag{1}$$

SLOPE:

$$\min_{b\in\mathbb{R}^p} \frac{1}{2}\|Y - Xb\|_2^2 + \sum_{i=1}^{p} \lambda_i |b|_{[i]}, \tag{2}$$

where $\lambda_1 \geq \ldots \geq \lambda_pand|b|_{[1]} \geq \ldots \geq |b|_{[p]}$.

Choosing λ sequence: How to select the λ sequence for metric control?

Methodology

Introduction: Lasso is commonly used for variable selection. In our study, we will explore variations of Lasso.

Lasso:

$$\min_{b \in \mathbb{R}^p} \frac{1}{2} \| Y - Xb \|_2^2 + \lambda \| b \|_1 \tag{1}$$

SLOPE:

$$\min_{b \in \mathbb{R}^p} \frac{1}{2} \| Y - Xb \|_2^2 + \sum_{i=1}^{p} \lambda_i |b|_{[i]}, \tag{2}$$

where $\lambda_1 \geq \ldots \geq \lambda_pand|b|_{[1]} \geq \ldots \geq |b|_{[p]}$.

Choosing λ sequence: How to select the λ sequence for metric control?

Intuitions: Orthogonal design matrix

Asumption: $X^T X = I_p$, σ known.
Our problem: Recover relevant covariates from β, where:

$$Y = X\beta + Z, \quad \text{where } Z \sim \mathcal{N}(0, \sigma^2 I_n)$$

We can see it as a multiple testing problem.

$$\widetilde{Y} \stackrel{\text{def}}{=} X^\top Y = X^\top X\beta + X^\top Z = \beta + X^\top Z \sim \mathcal{N}(\beta, \sigma^2 I_p)$$

How to solve it ?

1. Bonferroni
2. BH

Bonferroni

$$\mathscr{R}_{\mathrm{Bonf}} = \left\{ i; \quad \check{p}_i = 2\left(1 - \Phi\left(\frac{|\widetilde{Y}_i|}{\sigma}\right)\right) < \frac{\alpha}{p} \right\}$$

$$\text{or equivalently } \left\{ i; \Phi^{-1}\left(1 - \frac{\alpha}{2p}\right) < \frac{|\widetilde{Y}_i|}{\sigma} \right\},$$

where α is the confidence level, and Φ is the cumulative distribution function of a standard Gaussian, with $\widetilde{Y} = X^\top Y$.
This reminds us of the Lasso thresholding with

$$\lambda_{\mathrm{Bonf}} = \sigma\Phi^{-1}\left(1 - \frac{\alpha}{2p}\right)$$

$$\widehat{k}^{\mathrm{BH}} = \max\left\{k \in [\![0,p]\!]; \check{p}_{(k)} \leq \frac{\alpha k}{p}\right\}$$
$$= \max\left\{k \in [\![0,p]\!]; \frac{|\widetilde{Y}_{[k]}|}{\sigma} \geq \Phi^{-1}\left(1 - \frac{k\alpha}{2p}\right)\right\}$$

We can apply the same reasoning as before with the Lasso-Bonf with λ_{BH} written as follows:

$$\lambda_{\mathrm{BH}}(i) = \left(\sigma\Phi^{-1}\left(1 - \frac{i\alpha}{2p}\right)\right), \quad 1 \leq i \leq p$$

Theoretical guarantees of SLOPE

Theorem 1 (SLOPE's FDR control)

Under the linear assumption, orthogonal X, $Z \sim \mathcal{N}(0, \sigma^2 I_n)$ with σ known, then SLOPE verifies $\mathrm{FDR} \leq \alpha p_0/p$.

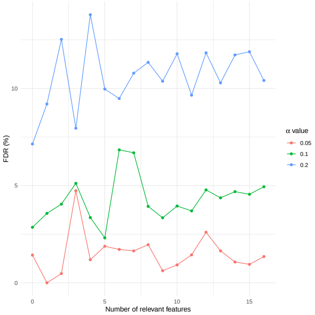


Figure 1: $n = 800, p = 800, \sigma^2 = 1, \alpha \in \{0.05, 0.1, 0.2\}$ and $|\mathcal{H}_1| \in [\![0;16]\!]$.

Relation with other methods

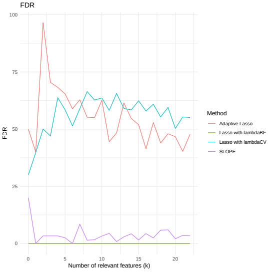


Figure 2: FDR

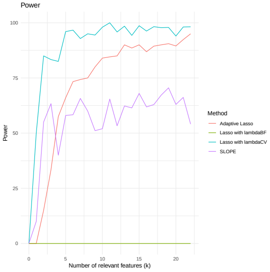


Figure 3: Power

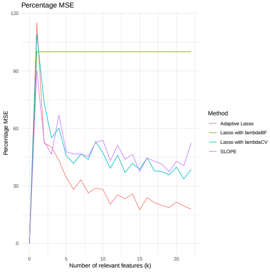


Figure 4: MSE

Figure 5: Different metrics to compare SLOPE, $\mathrm{Lasso} - \lambda_{\mathrm{Bonf}}$, $\mathrm{Lasso} - \lambda_{\mathrm{CV}}$ and $\mathrm{Adaptive} - \mathrm{Lasso}$ (with $\alpha = 0.1, n = p = 500, \sigma^2 = 1$).

Index

Intuitions to generalize SLOPE

⚠Shrinkage due to penalization.
Define $\eta_\lambda(t) = \mathrm{sign}(t)(|t| - \lambda)_+$ and assume that $X_j^\top X_j = 1$ and $\beta \geq 0$

1. Lasso optimality condition:

$$\widehat{\beta} = \eta_\lambda(\widehat{\beta} - X^\top(X\widehat{\beta} - Y)) = \eta_\lambda(\widehat{\beta} - X^\top X(\widehat{\beta} - \beta) + X^\top z).$$

2. Define $v_i := \langle X_i, \sum_{i \neq j} X_j(\beta_j - \widehat{\beta}_j)\rangle$ so that $\widehat{\beta}_i = \eta_\lambda(\beta_i + X_i^\top z + v_i)$.
💡Estimate v_i to iteratively adapt the penalization sequence.

Intuitions to generalize SLOPE

⚠ Shrinkage due to penalization.
Define $\eta_\lambda(t) = \mathrm{sign}(t)(|t| - \lambda)_+$ and assume that $X_j^\top X_j = 1$ and $\beta \geq 0$

1. Lasso optimality condition:

$$\widehat{\beta} = \eta_\lambda(\widehat{\beta} - X^\top(X\widehat{\beta} - Y)) = \eta_\lambda(\widehat{\beta} - X^\top X(\widehat{\beta} - \beta) + X^\top z).$$

2. Define $v_i := \langle X_i, \sum_{i \neq j} X_j(\beta_j - \widehat{\beta}_j)\rangle$ so that $\widehat{\beta}_i = \eta_\lambda(\beta_i + X_i^\top z + v_i)$.
   💡Estimate v_i to iteratively adapt the penalization sequence.

Intuitions to generalize SLOPE

⚠Shrinkage due to penalization.
Define $\eta_\lambda(t) = \mathrm{sign}(t)(|t| - \lambda)_+$ and assume that $X_j^\top X_j = 1$ and $\beta \geq 0$

1. Lasso optimality condition:

$$\widehat{\beta} = \eta_\lambda(\widehat{\beta} - X^\top(X\widehat{\beta} - Y)) = \eta_\lambda(\widehat{\beta} - X^\top X(\widehat{\beta} - \beta) + X^\top z).$$

2. Define $v_i := \langle X_i, \sum_{i \neq j} X_j(\beta_j - \widehat{\beta}_j)\rangle$ so that $\widehat{\beta}_i = \eta_\lambda(\beta_i + X_i^\top z + v_i)$.
   💡Estimate v_i to iteratively adapt the penalization sequence.

Intuitions to generalize SLOPE

3. KKT conditions on the Lasso give $X_{\mathscr{S}}^{\top}(Y - X\widehat{\beta}_{\mathscr{S}}) = \lambda 1_{\mathscr{S}}$, or equivalently, $\widehat{\beta}_{\mathscr{S}} = (X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}(X_{\mathscr{S}}^{\top}Y - \lambda 1_{\mathscr{S}})$
4. Adapting it to SLOPE: $\widehat{\beta}_{\mathscr{S}} = (X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}(X_{\mathscr{S}}^{\top}Y - \lambda_{\mathscr{S}})$
5. $\mathbb{E}[\beta_{\mathscr{S}} - \widehat{\beta}_{\mathscr{S}}] \approx (X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}}$, i.e. $\mathbb{E}[X_i^{\top}X_{\mathscr{S}}(\beta_{\mathscr{S}} - \widehat{\beta}_{\mathscr{S}})] \approx X_i^{\top}X_{\mathscr{S}}(X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}}$

Finally, if $X \sim \mathscr{N}(0, 1/n)$, we obtain that $\mathbb{E}[(X_i^{\top}X_{\mathscr{S}}(X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}})^2] = \lambda_{\mathscr{S}}^{\top}\mathbb{E}[(X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}]\lambda_{\mathscr{S}}/n = w(|\mathscr{S}|)\|\lambda_{\mathscr{S}}\|_2^2$ where $w(k) = 1/(n-k-1)$.

Intuitions to generalize SLOPE

3. KKT conditions on the Lasso give $X_{\mathscr{S}}^{\top}(Y - X\widehat{\beta}_{\mathscr{S}}) = \lambda 1_{\mathscr{S}}$, or equivalently, $\widehat{\beta}_{\mathscr{S}} = (X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}(X_{\mathscr{S}}^{\top}Y - \lambda 1_{\mathscr{S}})$
4. Adapting it to SLOPE: $\widehat{\beta}_{\mathscr{S}} = (X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}(X_{\mathscr{S}}^{\top}Y - \lambda_{\mathscr{S}})$
5. $\mathbb{E}[\beta_{\mathscr{S}} - \widehat{\beta}_{\mathscr{S}}] \approx (X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}}$, i.e. $\mathbb{E}[X_i^{\top}X_{\mathscr{S}}(\beta_{\mathscr{S}} - \widehat{\beta}_{\mathscr{S}})] \approx X_i^{\top}X_{\mathscr{S}}(X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}}$

Finally, if $X \sim \mathscr{N}(0, 1/n)$, we obtain that $\mathbb{E}[(X_i^{\top}X_{\mathscr{S}}(X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}})^2] = \lambda_{\mathscr{S}}^{\top}\mathbb{E}[(X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}]\lambda_{\mathscr{S}}/n = w(|\mathscr{S}|)\|\lambda_{\mathscr{S}}\|_2^2$ where $w(k) = 1/(n-k-1)$.

Intuitions to generalize SLOPE

3. KKT conditions on the Lasso give $X_{\mathscr{S}}^{\top}(Y - X\widehat{\beta}_{\mathscr{S}}) = \lambda 1_{\mathscr{S}}$, or equivalently, $\widehat{\beta}_{\mathscr{S}} = (X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}(X_{\mathscr{S}}^{\top}Y - \lambda 1_{\mathscr{S}})$
4. Adapting it to SLOPE: $\widehat{\beta}_{\mathscr{S}} = (X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}(X_{\mathscr{S}}^{\top}Y - \lambda_{\mathscr{S}})$
5. $\mathbb{E}[\beta_{\mathscr{S}} - \widehat{\beta}_{\mathscr{S}}] \approx (X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}}$, i.e.
$\mathbb{E}[X_i^{\top}X_{\mathscr{S}}(\beta_{\mathscr{S}} - \widehat{\beta}_{\mathscr{S}})] \approx X_i^{\top}X_{\mathscr{S}}(X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}}$

Finally, if $X \sim \mathscr{N}(0, 1/n)$, we obtain that
$\mathbb{E}[(X_i^{\top}X_{\mathscr{S}}(X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}})^2] = \lambda_{\mathscr{S}}^{\top}\mathbb{E}[(X_{\mathscr{S}}^{\top}X_{\mathscr{S}})^{-1}]\lambda_{\mathscr{S}}/n = w(|\mathscr{S}|)\|\lambda_{\mathscr{S}}\|_2^2$
where $w(k) = 1/(n - k - 1)$.

Intuitions to generalize SLOPE

3. KKT conditions on the Lasso give $X_{\mathscr{S}}^{\top}(Y - X\widehat{\beta}_{\mathscr{S}}) = \lambda 1_{\mathscr{S}}$, or equivalently, $\widehat{\beta}_{\mathscr{S}} = (X_{\mathscr{S}}^{\top} X_{\mathscr{S}})^{-1}(X_{\mathscr{S}}^{\top} Y - \lambda 1_{\mathscr{S}})$
4. Adapting it to SLOPE: $\widehat{\beta}_{\mathscr{S}} = (X_{\mathscr{S}}^{\top} X_{\mathscr{S}})^{-1}(X_{\mathscr{S}}^{\top} Y - \lambda_{\mathscr{S}})$
5. $\mathbb{E}[\beta_{\mathscr{S}} - \widehat{\beta}_{\mathscr{S}}] \approx (X_{\mathscr{S}}^{\top} X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}}$, i.e.
$\mathbb{E}[X_i^{\top} X_{\mathscr{S}}(\beta_{\mathscr{S}} - \widehat{\beta}_{\mathscr{S}})] \approx X_i^{\top} X_{\mathscr{S}}(X_{\mathscr{S}}^{\top} X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}}$

Finally, if $X \sim \mathscr{N}(0, 1/n)$, we obtain that
$\mathbb{E}[(X_i^{\top} X_{\mathscr{S}}(X_{\mathscr{S}}^{\top} X_{\mathscr{S}})^{-1}\lambda_{\mathscr{S}})^2] = \lambda_{\mathscr{S}}^{\top}\mathbb{E}[(X_{\mathscr{S}}^{\top} X_{\mathscr{S}})^{-1}]\lambda_{\mathscr{S}}/n = w(|\mathscr{S}|)\|\lambda_{\mathscr{S}}\|_2^2$
where $w(k) = 1/(n-k-1)$.

SLOPE with $\lambda_{G^\star}$

1. $\lambda_G(1) = \lambda_{BH}(1)$
2. $\lambda_G(2) = \lambda_{BH}(2)\sqrt{1 + w(1)\lambda_G(1)^2}$
3. $\lambda_G(i) = \lambda_{BH}(i)\sqrt{1 + w(i-1)\sum_{j<i}\lambda_G(j)^2}$

⚠ λ must be decreasing to guarantee convexity of the problem, then if there exist $i_0 \in \{0, \dots, p-1\}$ such that $\lambda_G(i_0) < \lambda_G(i_0+1)$, fix the rest the sequence to $\lambda_G(i_0)$, which gives $\lambda_{G^\star}$.

SLOPE with $\lambda_{G^\star}$

1. $\lambda_{\mathrm{G}}(1) = \lambda_{\mathrm{BH}}(1)$
2. $\lambda_{\mathrm{G}}(2) = \lambda_{\mathrm{BH}}(2)\sqrt{1 + w(1)\lambda_{\mathrm{G}}(1)^2}$
3. $\lambda_{\mathrm{G}}(i) = \lambda_{\mathrm{BH}}(i)\sqrt{1 + w(i-1)\sum_{j<i}\lambda_{\mathrm{G}}(j)^2}$

⚠ λ must be decreasing to guarantee convexity of the problem, then if there exist $i_0 \in \{0, \ldots, p-1\}$ such that $\lambda_{\mathrm{G}}(i_0) < \lambda_{\mathrm{G}}(i_0+1)$, fix the rest the sequence to $\lambda_{\mathrm{G}}(i_0)$, which gives $\lambda_{\mathrm{G}^\star}$.

SLOPE with $\lambda_{G^\star}$

1. $\lambda_{\mathrm{G}}(1) = \lambda_{\mathrm{BH}}(1)$
2. $\lambda_{\mathrm{G}}(2) = \lambda_{\mathrm{BH}}(2)\sqrt{1 + w(1)\lambda_{\mathrm{G}}(1)^2}$
3. $\lambda_{\mathrm{G}}(i) = \lambda_{\mathrm{BH}}(i)\sqrt{1 + w(i-1)\sum_{j<i}\lambda_{\mathrm{G}}(j)^2}$

⚠ λ must be decreasing to guarantee convexity of the problem, then if there exist $i_0 \in \{0, \dots, p-1\}$ such that $\lambda_{\mathrm{G}}(i_0) < \lambda_{\mathrm{G}}(i_0+1)$, fix the rest the sequence to $\lambda_{\mathrm{G}}(i_0)$, which gives $\lambda_{\mathrm{G}^\star}$.

SLOPE with $\lambda_{G^\star}$

1. $\lambda_G(1) = \lambda_{BH}(1)$
2. $\lambda_G(2) = \lambda_{BH}(2)\sqrt{1 + w(1)\lambda_G(1)^2}$
3. $\lambda_G(i) = \lambda_{BH}(i)\sqrt{1 + w(i-1)\sum_{j<i}\lambda_G(j)^2}$

⚠ λ must be decreasing to guarantee convexity of the problem, then if there exist $i_0 \in \{0, \dots, p-1\}$ such that $\lambda_G(i_0) < \lambda_G(i_0+1)$, fix the rest the sequence to $\lambda_G(i_0)$, which gives $\lambda_{G^\star}$.

Index

Is SLOPE better than BH?

Assume $y \sim \mathcal{N}(\mu, \Sigma)$ with Σ *almost* diagonal, $\Sigma_{i,j} \geq 0$ and known.

Goal:

What are the relevant covariates?(i.e. $\{i \in \{1, \dots p\}; \mu_i \neq 0\}$

- **BH:** Order $|y|_{[1]} \geq \dots \geq |y|_{[p]}$ and then SU with $\tau(i) = (\sigma\Phi^{-1}(1 - i\alpha/2p))_i$.
- **SLOPE:** $\underbrace{\tilde{y}}_{\text{Label}} = \Sigma^{-1/2} y = \underbrace{\Sigma^{-1/2}}_{\text{Design}} \mu + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, I_p)$ with $\lambda_{\mathrm{G}*}$.

Is SLOPE better than BH?

Assume $y \sim \mathcal{N}(\mu, \Sigma)$ with Σ *almost* diagonal, $\Sigma_{i,j} \geq 0$ and known.

Goal:

What are the relevant covariates?(i.e. $\{i \in \{1, \dots p\}; \mu_i \neq 0\}$

- **BH:** Order $|y|_{[1]} \geq \dots \geq |y|_{[p]}$ and then SU with $\tau(i) = (\sigma\Phi^{-1}(1 - i\alpha/2p))_i$.
- **SLOPE:** $\underbrace{\widetilde{y}}_{\text{Label}} = \Sigma^{-1/2} y = \underbrace{\Sigma^{-1/2}}_{\text{Design}} \mu + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, I_p)$ with λ_{G*}.

Is SLOPE better than BH?

Assume $y \sim \mathcal{N}(\mu, \Sigma)$ with Σ *almost* diagonal, $\Sigma_{i,j} \geq 0$ and known.

Goal:

What are the relevant covariates?(i.e. $\{i \in \{1, \dots p\}; \mu_i \neq 0\}$

- **BH:** Order $|y|_{[1]} \geq \dots \geq |y|_{[p]}$ and then SU with $\tau(i) = (\sigma\Phi^{-1}(1 - i\alpha/2p))_i$.
- **SLOPE:** $\underbrace{\widetilde{y}}_{\text{Label}} = \Sigma^{-1/2} y = \underbrace{\Sigma^{-1/2}}_{\text{Design}} \mu + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, I_p)$ with $\lambda_{\mathrm{G}*}$.

Is SLOPE better than BH?

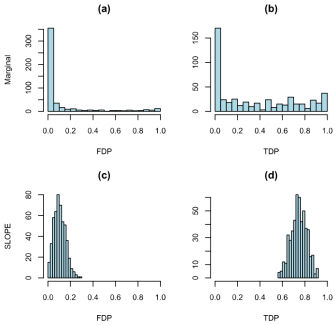


Figure 6: (From [Bogdan et al.(2015a)])The FDP of BH is either 0 or uniformly concentrated on $[0, 1]$, thus the discoveries are not trustworthy. ($\alpha = 0.1$)

Open question:

Can we get guarantees on the FDP instead of on the FDR for SLOPE?

Index

Conclusions and Perspectives

- ✓ Tractable convex optimization problem.
- ✓ Adaptability of BH.
- ✓ Control of the FDR(under strong assumptions).
- **?** Control of the FDP instead of just FDR?
- **?** SLOPE with $\lambda_{\mathrm{G}^\star}$, is it that general?
- × Reproducibility of the experiments from the article.

Index

References

[Abramovich and Benjamini(1995)] Felix Abramovich and Yoav Benjamini.
Thresholding of Wavelet Coefficients as Multiple Hypotheses Testing Procedure, pages 5–14.
Springer New York, New York, NY, 1995.
ISBN 978-1-4612-2544-7.
doi: 10.1007/978-1-4612-2544-7_1.
URL https://doi.org/10.1007/978-1-4612-2544-7_1.

[Benjamini and Hochberg(1995)] Yoav Benjamini and Yosef Hochberg.
Controlling the false discovery rate: A practical and powerful approach to multiple testing.
Journal of the Royal Statistical Society: Series B (Methodological), 57(1):289–300, 1995.
doi: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x.

[Benjamini and Yekutieli(2001)] Yoav Benjamini and Daniel Yekutieli.
The control of the false discovery rate in multiple testing under dependency.
The Annals of Statistics, 29(4):1165 – 1188, 2001.
doi: 10.1214/aos/1013699998.
URL https://doi.org/10.1214/aos/1013699998.

[Birgé and Massart(2001)] Lucien Birgé and Pascal Massart.
Gaussian model selection.
Journal of the European Mathematical Society, 3:203–268, 2001.
URL https://api.semanticscholar.org/CorpusID:54611071.

References

[Bogdan et al.(2015a)Bogdan, van den Berg, Sabatti, Su, and Candès]

Małgorzata Bogdan, Ewout van den Berg, Chiara Sabatti, Weijie Su, and Emmanuel J. Candès.
Slope—adaptive variable selection via convex optimization.
The Annals of Applied Statistics, 9(3), September 2015a.
ISSN 1932-6157.
doi: 10.1214/15-aoas842.
URL http://dx.doi.org/10.1214/15-AOAS842.

[Bogdan et al.(2015b)Bogdan, van den Berg, Sabatti, Su, and Candès]

Małgorzata Bogdan, Ewout van den Berg, Chiara Sabatti, Weijie Su, and Emmanuel J. Candès.
Supplement to "SLOPE—Adaptive variable selection via convex optimization".
https://doi.org/10.1214/15-AOAS842SUPP, 2015b.
URL https://doi.org/10.1214/15-AOAS842SUPP.

References

Online Appendix containing proofs of some technical results discussed in the text.

[Bonferroni(1936)] C. Bonferroni.
Teoria statistica delle classi e calcolo delle probabilita.
8:3–62, 1936.

[Giraud(2021)] Christophe Giraud.
Introduction to high-dimensional statistics.
CRC Press, 2021.

[Holm(1979)] Sture Holm.
A simple sequentially rejective multiple test procedure.
Scandinavian Journal of Statistics, 6(2):65–70, 1979.
ISSN 03036898, 14679469.
URL `http://www.jstor.org/stable/4615733`.

References

[Larsson et al.(2022)Larsson, Wallin, Bogdan, van den Berg, Sabatti, Car

Johan Larsson, Jonas Wallin, Malgorzata Bogdan, Ewout van den Berg, Chiara Sabatti, Emmanuel Candes, Evan Patterson, Weijie Su, Jakub Kała, Krystyna Grzesiak, and Michal Burdukiewicz.
SLOPE: Sorted L1 Penalized Estimation, 2022.
URL https://CRAN.R-project.org/package=SLOPE.
R package version 0.5.0.

[Lederer(2013)] Johannes Lederer.
Trust, but verify: benefits and pitfalls of least-squares refitting in high dimensions, 2013.

References

[Tay et al.(2023)Tay, Narasimhan, and Hastie] J. Kenneth Tay, Balasubramanian Narasimhan, and Trevor Hastie. Elastic net regularization paths for all generalized linear models. *Journal of Statistical Software*, 106(1):1–31, 2023. doi: 10.18637/jss.v106.i01.

[Zou(2006)] Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476): 1418–1429, 2006. doi: 10.1198/016214506000000735. URL https://doi.org/10.1198/016214506000000735.

Thank You, Questions?

Index

Article figures: FDR not controlled by SLOPE-BH

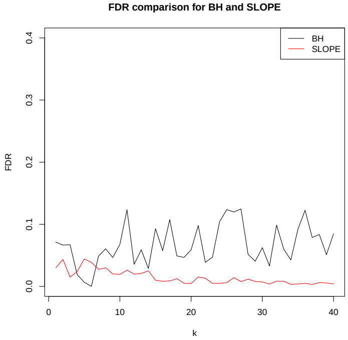


Figure 7: FDR controlled for the experience of BH vs SLOPE. However, SLOPE looks too restrictive...

Article figures: FDR not controlled by SLOPE-BH

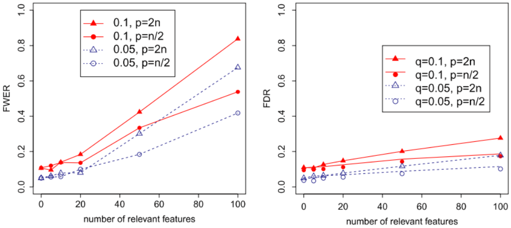


Figure 8: (From [Bogdan et al.(2015a)]) On the left FWER for Lasso-λ_{Bonf} and on the right the FDR for SLOPE with λ_{BH} under Gaussian correlated design.

Article figures: Relation with other methods

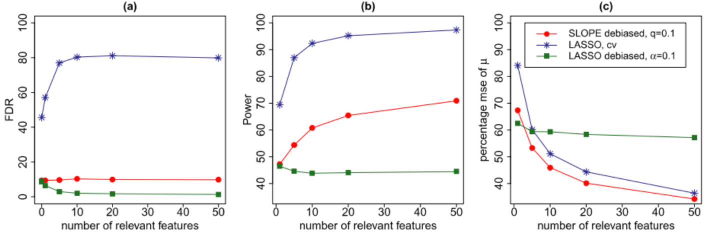


Figure 9: (From [Bogdan et al.(2015a)])SLOPE-$\lambda_{G^\star}$ vs Lasso-λ_{CV} and Lasso-λ_{Bonf} with design entries i.i.d. $\mathcal{N}(0,1/n)$, $n = p = 5000, \sigma^2 = 1$ and non zeros coefficients of the order of $2\sqrt{\log p}$.

Article figures: BH vs SLOPE

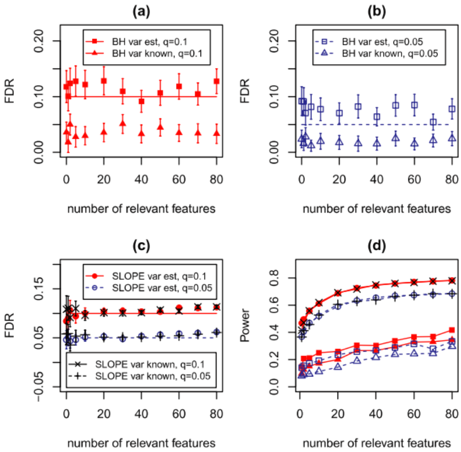


Figure 10: (From [Bogdan et al.(2015a)])Experiences to compare SLOPE-$\lambda_{G^\star}$ with BH procedure, with either known or unknown variance.