Supplementary Material: Appendix

The appendix is organized as follows. In section A we prove some properties of J_λ norm, which are useful for the article. In section B we prove the convexity of the objective of group SLOPE. Dual norm and the stopping criterion for the group SLOPE algorithm are discussed in section C. Section D presents the formulation of the group SLOPE in case when different groups are orthogonal to each other. Section E is concerned with the properties of SLOPE when the design matrix is positive on the diagonal and has zeros at off-diagonal entries. These properties are crucial for the Theorem about gFDR control of gSLOPE (Theorem 2.5). The Theorem 2.6 on the minimax properties of SLOPE is proven in Section F, while Section G describes in detail the heuristic procedure for deriving the λ sequence when variables from different groups are independent. Sections H focuses on the calculation of the expected maximal group "effect" under the total null hypothesis, while the screening procedure used to obtain SNPs for the simulation study is described in Appendix I.

A J_λ norm properties

For nonnegative, nonincreasing sequence $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$ consider function $\mathbb{R}^p \ni b \longmapsto J_\lambda(b) \in \mathbb{R}$ given by $J_\lambda(b) = \sum_{i=1}^p \lambda_i \cdot |b|_{(i)}$, where $|b|_{(1)} \geq \ldots \geq |b|_{(p)}$ is the vector of sorted absolute values.

Proposition A.1. *If a, $b \in \mathbb{R}^p$ are such that $|a| \preceq |b|$, then $|a|_{(\cdot)} \preceq |b|_{(\cdot)}$.*

Proof. Without loss of generality we can assume that a and b are nonnegative and that it occurs $a_1 \geq \ldots \geq a_p$. We will show that $a_k \leq b_{(k)}$ for $k \in \{1, \ldots, p\}$. Fix such k and consider the set $S_k := \{b_i : \ b_i \geq a_k\}$. It is enough to show that $|S_k| \geq k$. For each $j \in \{1, \ldots, k\}$ we have

$$b_j \geq a_j \geq a_k \implies b_j \in S_k,$$

what proves the last statement. ■

Corollary A.2. *Let $a \in \mathbb{R}^p$, $b \in \mathbb{R}^p$ and $|a| \preceq |b|$ then Proposition (A.1) instantly gives that $J_\lambda(a) \leq J_\lambda(b)$, since $J_\lambda(a) = \lambda^\mathsf{T}|a|_{(\cdot)} \leq \lambda^\mathsf{T}|b|_{(\cdot)} = J_\lambda(b)$.*

Proposition A.3. *For fixed sequence $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$, let $b \in \mathbb{R}^p$ be such that $b \succeq 0$ and $b_j > b_l$ for some $j, l \in \{1, \ldots, p\}$. For $0 < \varepsilon \leq (b_j - b_l)/2$, define $b_\varepsilon \in \mathbb{R}^p$ by conditions $(b_\varepsilon)_l := b_l + \varepsilon$, $(b_\varepsilon)_j := b_j - \varepsilon$ and $(b_\varepsilon)_i := b_i$ for $i \notin \{j, l\}$. Then $J_\lambda(b_\varepsilon) \leq J_\lambda(b)$.*

Proof. Let $\pi : \{1, \ldots, p\} \longrightarrow \{1, \ldots, p\}$ be permutation such as $\sum_{i=1}^p \lambda_i (b_\varepsilon)_{(i)} = \sum_{i=1}^p \lambda_{\pi(i)} (b_\varepsilon)_i$ for each i in $\{1, \ldots, p\}$ and $\lambda_{\pi(j)} \geq \lambda_{\pi(l)}$. From the rearrangement inequality (Theorem 368 in Hardy et al., 1952),

$$\begin{aligned} J_\lambda(b) - J_\lambda(b_\varepsilon) = & \sum_{i=1}^p \lambda_i b_{(i)} - \sum_{i=1}^p \lambda_i (b_\varepsilon)_{(i)} = \sum_{i=1}^p \lambda_i b_{(i)} - \sum_{i=1}^p \lambda_{\pi(i)} (b_\varepsilon)_i \\ \geq & \sum_{i=1}^p \lambda_{\pi(i)} b_i - \sum_{i=1}^p \lambda_{\pi(i)} (b_\varepsilon)_i = \varepsilon\big(\lambda_{\pi(j)} - \lambda_{\pi(l)}\big) \geq 0. \end{aligned} \tag{A.1}$$

■

B Convexity of the objective function

To show that the objectives in problems (2.2) and (2.4) are convex functions, we will prove the following propositions

Proposition B.1. *Function $J_{\lambda,\mathbb{I},W}(b) := J_\lambda\Big(W[\![b]\!]_\mathbb{I}\Big)$ is a norm for any nonnegative, nonincreasing sequence $\{\lambda_i\}_{i=1}^m$ containing at least one nonzero element, partition $\mathbb{I}$ of the set $\{1, \ldots, \widetilde{p}\}$ and diagonal matrix W with positive elements on diagonal.*

Proof. It is easy to see that $J_{\lambda,\mathbb{I},W}(c) = 0$ if and only if $c = 0$ and that for any scalar $\alpha \in \mathbb{R}$ it occurs $J_{\lambda,\mathbb{I},W}(\alpha c) = |\alpha| J_{\lambda,\mathbb{I},W}(c)$. We will show that $J_{\lambda,\mathbb{I},W}$ satisfies the triangle inequality. Let b, c be any vectors from $\mathbb{R}^{\widetilde{p}}$. From the positivity of w_i's we have $W[\![a+b]\!]_{\mathbb{I}} \preceq W[\![a]\!]_{\mathbb{I}} + W[\![b]\!]_{\mathbb{I}}$. Therefore, Corollary A.2 yields

$$\begin{aligned} J_{\lambda,\mathbb{I},W}\big(a+b\big) = \ & J_\lambda\Big(W[\![a+b]\!]_{\mathbb{I}}\Big) \leq \ J_\lambda\Big(W[\![a]\!]_{\mathbb{I}} + W[\![b]\!]_{\mathbb{I}}\Big) \\ & \leq \ J_\lambda\Big(W[\![a]\!]_{\mathbb{I}}\Big) + J_\lambda\Big(W[\![b]\!]_{\mathbb{I}}\Big) = J_{\lambda,\mathbb{I},W}(a) + J_{\lambda,\mathbb{I},W}(b), \end{aligned} \tag{B.1}$$

since J_λ is a norm. ■

Proposition B.2. *Function $J_\lambda\Big(W[\![b]\!]_{I,X}\Big)$ is a seminorm for any nonnegative, nonincreasing sequence $\{\lambda_i\}_{i=1}^m$, partition I of the set $\{1,\ldots,p\}$, design matrix $X \in M(n,p)$ and diagonal matrix W with positive elements on diagonal.*

Proof. Clearly, $J_\lambda\Big(W[\![\alpha b]\!]_{I,X}\Big) = |\alpha| J_\lambda\Big(W[\![b]\!]_{I,X}\Big)$, for any scalar $\alpha \in \mathbb{R}$. Moreover, for any $a, b \in \mathbb{R}^p$, it holds $W[\![a+b]\!]_{I,X} \preceq W[\![a]\!]_{I,X} + W[\![b]\!]_{I,X}$, and the triangle inequality could be proved similarly as in the previous proposition. ■

C Stopping criterion

C.1 Dual norm and conjugate of grouped sorted ℓ_1 norm

Let $f : \mathbb{R}^p \to \mathbb{R}$ be a norm. We will use notation f^D to refer to the dual norm to f, i.e function defined as $f^D(x) := \max_b \big\{x^\mathsf{T} b \,:\, f(b) \leq 1\big\}$. It could be shown (see Bogdan et al., 2015), that the set C_λ, defined as $C_\lambda := \Big\{x \in \mathbb{R}^p : \ \sum_{i=1}^k |x|_{(i)} \leq \sum_{i=1}^k \lambda_i, \ k = 1,\ldots,p\Big\}$, is unit ball of the dual norm to J_λ for any nonnegative, nonincreasing sequence $\{\lambda_i\}_{i=1}^p$ with at least one nonzero element. We will now consider the dual norm to $J_{\lambda,I,W}(b) = J_\lambda\big(W[\![b]\!]_I\big)$. It holds

$$\begin{aligned} J^D_{\lambda,I,W}(x) = \max_b\big\{x^\mathsf{T} b : \ J_{\lambda,I,W}(b) \leq 1\big\} = \max_b\big\{x^\mathsf{T} b : \ J_\lambda(W[\![b]\!]_I) \leq 1\big\} = \\ \max_{b,c}\big\{x^\mathsf{T} b : \ J_\lambda(c) \leq 1, \ c = W[\![b]\!]_I\big\} = \max_c\big\{x^\mathsf{T} b^c : \ J_\lambda(c) \leq 1, \ c \succeq 0\big\}, \end{aligned} \tag{C.1}$$

where b^c is defined as $b^c := \operatorname{argmax}_b \left\{x^\top b : \ c = W[\![b]\!]_I\right\}$. This problem is separable and for each i we have $b^c_{I_i} = \operatorname{argmax}\left\{x_{I_i}^\top b_{I_i} : \ c_i^2 = w_i^2\|b_{I_i}\|_2^2\right\}$. Applying the Lagrange multiplier method quickly yields $x_{I_i}^\top b^c_{I_i} = c_i w_i^{-1}\|x_{I_i}\|_2$. Consequently,

$$\begin{aligned} J^D_{\lambda,I,W}(x) = \max_c \big\{(W^{-1}[\![x]\!]_I)^\top c : \ J_\lambda(c) \le 1, \ c \succeq 0\big\} = \\ \max_c \big\{(W^{-1}[\![x]\!]_I)^\top c : \ J_\lambda(c) \le 1\big\} = \ J^D_\lambda\big(W^{-1}[\![x]\!]_I\big). \end{aligned} \tag{C.2}$$

Therefore, $\left\{x : \ J^D_{\lambda,I,W}(x) \le 1\right\} = \left\{x : \ J^D_\lambda(W^{-1}[\![x]\!]_I) \le 1\right\} = \left\{x : \ W^{-1}[\![x]\!]_I \in C_\lambda\right\}$. Since the conjugate of a norm is equal to zero for arguments from the unit ball of the dual norm, and equal to infinity otherwise, we immediately get

Corollary C.1. *The conjugate function for $J_{\lambda,I,W}$ is the function $J^*_{\lambda,I,W}$ defined as*

$$J^*_{\lambda,I,W}(x) = \begin{cases} 0, \ W^{-1}[\![x]\!]_I \in C_\lambda \\ \infty, \ otherwise \end{cases}. \tag{C.3}$$

C.2 Stopping criteria for numerical algorithm

Without loss of generality assume that $\sigma = 1$. We will start with optimization problem presented in subsection 2.2, namely

$$\underset{\eta}{\text{minimize}} \quad f(\eta) = \frac{1}{2}\|y - \widetilde{X}M\eta\|_2^2 + J_\lambda\big([\![\eta]\!]_{\mathbb{I}}\big) \tag{C.4}$$

for $[\![\eta]\!]_{\mathbb{I}} = \Big(\|\eta_{\mathbb{I}_1}\|_2, \ldots, \|\eta_{\mathbb{I}_m}\|_2\Big)^\top$ and $M_{\mathbb{I}_i,\mathbb{I}_i} = \frac{1}{w_i}\mathbf{I}_{l_i}$, $i = 1, \ldots, m$. This problem could be written in equivalent form

$$\begin{aligned} &\underset{\eta,r,c}{\text{minimize}} \ \tfrac{1}{2}\|r\|_2^2 + c \\ &\text{s.t.} \quad \begin{cases} J_{\lambda,\mathbb{I}}(\eta) - c \le 0 \\ y - r - \widetilde{X}M\eta = 0 \end{cases} \end{aligned} \tag{C.5}$$

$\big($notice that for (η^*, r^*, c^*) to be a solution, it must be that $c^* = J_{\lambda,\mathbb{I}}(\eta^*)\big)$. Since $(C.5)$ is convex and (η_0, r_0, c_0), for $\eta_0 = 0$, $r_0 = y$ and $c_0 = 1$, is strictly feasible, the strong duality holds. The Lagrange dual function for this problem is given by

$$g(\mu,\nu) = \inf_{\eta,r,c} \left\{ \frac{1}{2}\|r\|_2^2 + c + \mu^\mathsf{T}\big(y - r - \widetilde{X}M\eta\big) + \nu\big(J_{\lambda,\mathbb{I}}(\eta) - c\big) \right\} =$$

$$\mu^\mathsf{T} y + \inf_{r} \left\{ \frac{1}{2}\|r\|_2^2 - \mu^\mathsf{T} r \right\} + \inf_{c} \big\{c - \nu c\big\} + \inf_{\eta} \big\{ -\mu^\mathsf{T}\widetilde{X}M\eta + \nu J_{\lambda,\mathbb{I}}(\eta)\big\}.$$

Now, since the minimum of $\frac{1}{2}\|r\|_2^2 - \mu^\mathsf{T} r$ is taken for $r = \mu$, we have

$$g(\mu,\nu) = \mu^\mathsf{T} y - \frac{1}{2}\|\mu\|_2^2 + \inf_{c} \big\{c - \nu c\big\} - J^*_{\nu\lambda,\mathbb{I}}\big((\widetilde{X}M)^\mathsf{T}\mu\big). \tag{C.6}$$

Then $\nu^* = 1$ and from Corollary C.1, the dual problem to (C.5) is equivalent to

$$\begin{array}{ll} \underset{\mu}{\text{maximize}} & \mu^\mathsf{T} y - \frac{1}{2}\|\mu\|_2^2 \\ \text{s.t.} & [\![M\widetilde{X}^\mathsf{T}\mu]\!]_{\mathbb{I}} \in C_\lambda \end{array}. \tag{C.7}$$

Let (η^*, r^*, c^*) be primal and $(\mu^*, 1)$ be dual solution to (C.5). Obviously, $\mu^* = r^* = y - \widetilde{X}M\eta^*$ and $c^* = J_{\lambda,\mathbb{I}}(\eta^*)$. Furthermore, from strong duality we have

$$\frac{1}{2}\|y - \widetilde{X}M\eta^*\|_2^2 + J_{\lambda,\mathbb{I}}(\eta^*) = (y - \widetilde{X}M\eta^*)^\mathsf{T} y - \frac{1}{2}\|y - \widetilde{X}M\eta^*\|_2^2, \tag{C.8}$$

which gives $(\widetilde{X}M\eta^*)^\mathsf{T}(y - \widetilde{X}M\eta^*) = J_{\lambda,\mathbb{I}}\big(\eta^*\big)$. Now, for current approximate $\eta^{[k]}$ of solution to (C.4), achieved after applying proximal gradient method, we define the current duality gap for k step as

$$\rho(\eta^{[k]}) = (\widetilde{X}M\eta^{[k]})^\mathsf{T}(y - \widetilde{X}M\eta^{[k]}) - J_{\lambda,\mathbb{I}}\big(\eta^{[k]}\big) \tag{C.9}$$

and we will determine the infeasibility of $\mu^{[k]} := y - \widetilde{X}M\eta^{[k]}$ by using the measure

$$\text{infeas}\big(\mu^{[k]}\big) := \max\Big\{ J^D_{\lambda,\mathbb{I}}\big(M\widetilde{X}^\mathsf{T}\mu^{[k]}\big) - 1, 0\Big\} \tag{C.10}$$

To define the stopping criteria we have applied the widely used procedure: treat $\rho(\eta^{[k]})$ as indicator telling how far $\eta^{[k]}$ is from true solution and terminate the algorithm when this difference and infeasibility measure are sufficiently small. Summarizing, we have derived algorithm according to scheme

Procedure 1 group SLOPE

input: infeas.tol: *positive number determining the tolerance for infeasibility*;
dual.tol: *positive number determining the tolerance for duality gap*;
$k := 0,\ \eta^{[0]},\ \mu^{[0]} := \mu(\eta^{[0]}),\ \text{infeas}^{[0]} := \text{infeas}\big(\mu^{[0]}\big),\ \rho^{[0]} := \rho(\eta^{[0]})$;
while ($\text{infeas}^{[k]} > \text{infeas.tol}$ or $\rho^{[k]} > \text{dual.tol}$) **do**
1. $k \leftarrow k+1$;
2. get $\eta^{[k]}$ from Procedure 1;
3. $\mu^{[k]} := \mu(\eta^{[k]})$;
4. $\text{infeas}^{[k]} := \text{infeas}\big(\mu^{[k]}\big),\ \rho^{[k]} := \rho(\eta^{[k]})$;

end while
$\beta_{gS} := M\eta^{[k]}$.

D Alternative representation in the orthogonal case

Suppose that the experiment matrix is orthogonal at group level, i.e. it holds $X_{I_i}^\mathsf{T} X_{I_j} = \mathbf{0}$, for every $i, j \in \{1, \dots, m\}$, $i \neq j$. In such a case, $\widetilde{X}$ in problem (2.4) is orthogonal matrix, i.e. $\widetilde{X}^\mathsf{T}\widetilde{X} = \mathbf{I}_{\widetilde{p}}$. If $n = \widetilde{p}$, i.e. $\widetilde{X}$ is a square and orthogonal matrix, we also have $\widetilde{X}\widetilde{X}^\mathsf{T} = \mathbf{I}_{\widetilde{p}}$ and it obeys $\|\widetilde{X}^\mathsf{T} b\|_2^2 = b^\mathsf{T}\widetilde{X}\widetilde{X}^\mathsf{T} b = \|b\|_2^2$ for $b \in \mathbb{R}^{\widetilde{p}}$. For the general case with $n \geq \widetilde{p}$, we can extend $\widetilde{X}$ to a square matrix by adding new orthonormal columns and defining $\widetilde{X}_C := \big[\widetilde{X}\ C\big]$, where C is composed of vectors (columns) being some complement to orthogonal basis of $\mathbb{R}^{\widetilde{p}}$. For $y \in \mathbb{R}^n$ and $b \in \mathbb{R}^{\widetilde{p}}$ we get:

$$\Big\|y - \widetilde{X}b\Big\|_2^2 = \Big\|\widetilde{X}_C^\mathsf{T}\Big(y - \widetilde{X}b\Big)\Big\|_2^2 = \left\|\begin{bmatrix}\widetilde{X}^\mathsf{T}\\ C^\mathsf{T}\end{bmatrix} y - \begin{bmatrix} b \\ \mathbf{0}\end{bmatrix}\right\|_2^2 = \Big\|\widetilde{X}^\mathsf{T} y - b\Big\|_2^2 + const, \quad \text{(D.1)}$$

which implies that under orthogonal situation the optimization problem in (2.4) could be recast as

$$\underset{b}{\operatorname{argmin}} \quad \left\{\frac{1}{2}\big\|\widetilde{y} - b\big\|_2^2 + \sigma J_\lambda\big(W[\![b]\!]_{\mathbb{I}}\big)\right\}, \quad \text{(D.2)}$$

for $\widetilde{y} := \widetilde{X}^\mathsf{T} y$. After introducing new variable to problem (D.2), namely $c \in \mathbb{R}^m$, we get the equivalent formulation

$$\underset{b,c}{\operatorname{argmin}} \quad \left\{\frac{1}{2}\big\|\widetilde{y} - b\big\|_2^2 + \sigma J_\lambda(c) :\ c = W[\![b]\!]_{\mathbb{I}}\right\}. \quad \text{(D.3)}$$

Proposition D.1. *Let* $f(b, c) : \mathbb{R}^p \times \mathbb{R}^m \longrightarrow \mathbb{R}$ *be any function and consider optimization problem* $\operatorname{argmin}_{b,c}\big\{f(b,c) :\ (b,c) \in \mathcal{D}\big\}$ *with unique solution* (b^*, c^*) *and feasible* *set $\mathcal{D} \subset \mathbb{R}^p \times \mathbb{R}^m$. Define $\mathcal{D}^c := \big\{c \in \mathbb{R}^m \mid \exists b \in \mathbb{R}^p : (b,c) \in \mathcal{D}\big\}$. Suppose that for any $c \in \mathcal{D}^c$, there exists unique solution, b^c, to problem* $\operatorname{argmin}_b \big\{f(b,c) :\ (b,c) \in \mathcal{D}\big\}$. *Moreover, assume that the solution to* $\operatorname{argmin}_c \big\{f(b^c,c) :\ c \in \mathcal{D}^c\big\}$ *is unique. Then, it occurs*

$$\begin{cases} c^* = \operatorname{argmin}_c \big\{f(b^c,c) :\ c \in \mathcal{D}^c\big\} \\ b^* = b^{c^*} \end{cases}. \tag{D.4}$$

Proof. Suppose that there exists $(b^0,c^0) \in \mathcal{D}$, such that $f(b^0,c^0) < f(b^*,c^*)$, where b^* and c^* are defined as in (D.4). We have

$$f(b^{c^0},c^0) \le f(b^0,c^0) < f(b^*,c^*) = f(b^{c^*},c^*), \tag{D.5}$$

which leads to the contradiction with definition of c^*. ■

We will apply the above proposition to (D.3). Let (b^*,c^*) be solution to (D.3). Then b^* is also solution to convex problem (D.2) with strictly convex objective function and therefore is unique. Since $c^* = W[\![b^*]\!]_{\mathbb{I}}$, c^* is unique as well. In considered situation $\mathcal{D}^c = \big\{c :\ c \succeq 0\big\}$. We will start with solving the problem $b^c = \operatorname{argmin}_b \left\{\frac{1}{2}\big\|\widetilde{y} - b\big\|_2^2 + \sigma J_\lambda(c) :\ c = W[\![b]\!]_{\mathbb{I}}\right\}$. The additive constant in the objective could be omitted. Moreover, for each $i \in \{1,\ldots,m\}$ we have

$$b^c_{\mathbb{I}_i} = \operatorname*{argmin}_{b_{\mathbb{I}_i}} \left\{\big\|\widetilde{y}_{\mathbb{I}_i} - b_{\mathbb{I}_i}\big\|_2^2 :\ w_i^2\|b_{\mathbb{I}_i}\|_2^2 - c_i^2 = 0\right\}. \tag{D.6}$$

The Lagrange Multipliers method quickly yields $b^c_{\mathbb{I}_i} = (w_i\|\widetilde{y}_{\mathbb{I}_i}\|_2)^{-1}c_i\widetilde{y}_{\mathbb{I}_i}$ and, consequently, it holds $\big\|\widetilde{y}_{\mathbb{I}_i} - b^c_{\mathbb{I}_i}\big\|_2^2 = \big(\|\widetilde{y}_{\mathbb{I}_i}\|_2 - w_i^{-1}c_i\big)^2$. From Proposition D.1, we get the following procedure for solution, b^*, to problem (D.2)

$$\begin{cases} c^* = \operatorname{argmin}_c \left\{\frac{1}{2}\sum_{i=1}^m \big(\|\widetilde{y}_{\mathbb{I}_i}\|_2 - w_i^{-1}c_i\big)^2 + J_{\sigma\lambda}(c)\right\} \\ b^*_{\mathbb{I}_i} = c_i^*\big(w_i\|\widetilde{y}_{\mathbb{I}_i}\|_2\big)^{-1}\widetilde{y}_{\mathbb{I}_i}, \quad i = 1,\ldots,m \end{cases} \tag{D.7}$$

(notice that we applied Proposition E.2 to omit the constraints $c \succeq 0$ and that the objective function in definition of c^* is strictly feasible, which guarantees the unique solution. The above procedure yields conclusion, that indices of groups estimated by gSLOPE as relevant coincide with the support of solution to SLOPE problem with diagonal matrix having inverses of weights $w_1, \ldots, w_m$ on diagonal. Moreover, after defining $\widetilde{\beta} \in \mathbb{R}^{\widetilde{p}}$ by conditions $\widetilde{\beta}_{\mathbb{I}_i} := R_i \beta_{I_i}$, $i = 1, \ldots, m$, we simply have $[\![\widetilde{\beta}]\!]_{\mathbb{I}} = [\![\beta]\!]_{I,X}$ and

$$\widetilde{y} = \widetilde{X}^T y = \widetilde{X}^T \bigg(\sum_{i=1}^m U_i R_i \beta_{I_i} + z \bigg) = \widetilde{X}^T \big(\widetilde{X} \widetilde{\beta} + z \big) = \widetilde{\beta} + \widetilde{X}^T z, \quad \text{hence } \widetilde{y} \sim \mathcal{N}\big(\widetilde{\beta},\ \sigma^2 \mathbf{I}_{\widetilde{p}}\big).$$

Summarizing, if the assumption about the orthogonality at groups level is in use, one can consider the statistically equivalent model $\widetilde{y} \sim \mathcal{N}\big(\widetilde{\beta},\ \sigma^2 \mathbf{I}_{\widetilde{p}}\big)$, define truly relevant groups via the support of $[\![\widetilde{\beta}]\!]_{\mathbb{I}}$ and treat the vector $[\![b^*]\!]_{\mathbb{I}} = \big(\frac{c_1^*}{w_1}, \ldots, \frac{c_m^*}{w_m}\big)$ as an gSLOPE estimate of group effect sizes, where b^* and c^* are defined in (2.12), i.e. it holds $[\![b^*]\!]_{\mathbb{I}} = [\![\beta^{\text{gS}}]\!]_{I,X}$ for any solution β^{gS} to problem (2.2).

E SLOPE with diagonal experiment matrix

In this section we investigate the SLOPE properties in the situation when the design matrix is square and has only diagonal entries, which are all positive. The key results are two final Propositions: E.5 and E.6 which are used in the proof of Theorem 2.5. The remaining content of this section is needed to prove these two crucial results.

Let $y \in \mathbb{R}^p$ be fixed vector and $d_1, \ldots, d_p$ be positive numbers. We will use notation $diag(d_1, \ldots, d_p)$ to define the diagonal matrix D such as $D_{i,i} = d_i$ for $i = 1, \ldots, p$. Denote $d := (d_1, \ldots, d_p)^\mathsf{T}$ and let b^* be the solution to SLOPE optimization problem with diagonal experiment matrix, i.e. the solution to

$$\underset{b}{\text{argmin}}\ f(b) := \Big\{ \frac{1}{2} \big\| y - Db \big\|_2^2 + J_\lambda\big(b\big) \Big\}. \tag{E.1}$$

Since f is strictly convex function, the solution to (E.1) is unique. It is easy to observe, that changing sign of y_i corresponds to changing sign at i^{th} coefficient of solution as well as permuting coefficients of y together with $d_i' s$ permutes coefficients of b^*. We will summarize this observation below without proofs.

Proposition E.1. *Let $\pi : \{1,\dots,p\} \longrightarrow \{1,\dots,p\}$ be given permutation with P_π as corresponding matrix. Then:*

- ***A.1*** $P_\pi D P_\pi^\intercal = diag(d_{\pi(1)},\dots,d_{\pi(p)})$,
- ***A.2*** $b_\pi := P_\pi b^*$ *is solution to* $\underset{b}{\text{minimize}}\ f_\pi(b) := \frac{1}{2}\left\|P_\pi y - P_\pi D P_\pi^\intercal b\right\|_2^2 + J_\lambda(b)$,
- ***A.3*** $b_S := S b^*$ *is solution to* $\underset{b}{\text{minimize}}\ f_S(b) := \frac{1}{2}\left\|Sy - Db\right\|_2^2 + J_\lambda(b)$,

  where S is diagonal matrix with entries on diagonal coming from set $\{-1,1\}$.

Proposition E.2. *If $y \succeq 0$, then $b^* \succeq 0$.*

Proof. Suppose that for some r it occurs $b_r < 0$ for any $b \in \mathbb{R}^p$. If $y_r = 0$, then taking $\widehat{b}$ defined as $\widehat{b}_i := \begin{cases} 0, & i = r \\ b_i, & \text{otherwise} \end{cases}$, we get $|\widehat{b}| \preceq |b|$ and Corollary A.2 gives $J_\lambda(\widehat{b}) \leq J_\lambda(b)$. Consequently,

$$f(b) - f(\widehat{b}) \geq \frac{1}{2}\left\|y - Db\right\|_2^2 - \frac{1}{2}\left\|y - D\widehat{b}\right\|_2^2 = \frac{1}{2}(y_r - d_r b_r)^2 - \frac{1}{2}(y_r + d_r \widehat{b}_r)^2 = \frac{1}{2}d_r^2 b_r^2 > 0.$$

Hence b could not be the solution. Now consider case when $y_r > 0$ and define $\widehat{b}$ by putting $\widehat{b}_i := \begin{cases} -b_r, & i = r \\ b_i, & \text{otherwise} \end{cases}$. Then we have $J_\lambda(b) = J_\lambda(\widehat{b})$ and

$$f(b) - f(\widehat{b}) = \frac{1}{2}(y_r - d_r b_r)^2 - \frac{1}{2}(y_r + d_r b_r)^2 = -2y_r d_r b_r > 0.$$

and, as before, b could not be optimal. ■

Proposition E.3. *Let b^* be the solution to problem (E.1), $\{y_i\}_{i=1}^p$ be nonnegative sequence, $\{d_i\}_{i=1}^p$ be the sequence of positive numbers and assume that*

$$d_1 y_1 \geq \ldots \geq d_p y_p. \tag{E.2}$$

If b^* has exactly r nonzero entries for $r > 0$, then the set $\{1,\dots,r\}$ corresponds to the support of b^*.

Proof. It is enough to show that

$$\big(j \in \{2,\dots,m\},\quad b_j^* \neq 0\big) \implies b_{j-1}^* \neq 0. \tag{E.3}$$

Suppose that this is not true. From Proposition E.2 we know that b^* is nonnegative, hence we can find i from $\{2,\dots,m\}$ such as $b_j^* > 0$ and $b_{j-1}^* = 0$. For $\varepsilon \in \big(0, b_j^*/2\big]$ define vector b_ε by putting $(b_\varepsilon)_{j-1} := \varepsilon$, $(b_\varepsilon)_j := b_j^* - \varepsilon$ and $(b_\varepsilon)_i := b_i^*$ for $i \notin \{j,l\}$. From Proposition A.3 we have that $J_\lambda(b_\varepsilon) \le J_\lambda(b^*)$, which gives

$$\begin{aligned} f(b^*) - f(b_\varepsilon) \ge \frac{1}{2}\big(y_{j-1} - d_{j-1}b_{j-1}^*\big)^2 + \frac{1}{2}\big(y_j - d_j b_j^*\big)^2 \\ - \frac{1}{2}\big(y_{j-1} - d_{j-1}b_\varepsilon(j-1)\big)^2 - \frac{1}{2}\big(y_j - d_j b_\varepsilon(j)\big)^2 = \\ \varepsilon\left(A - \frac{d_{j-1}^2 + d_j^2}{2}\cdot\varepsilon\right), \\ \text{for } A := (y_{j-1}d_{j-1} - y_j d_j) + d_j^2 b_j^* > 0. \end{aligned} \tag{E.4}$$

Therefore, $f(b^*) > f(b_\varepsilon)$ for some $\varepsilon > 0$, which contradicts the optimality of b^*. ■

Consider now problem (E.1) with arbitrary sequence $\{y_i\}_{i=1}^p$. Suppose that b^* has exactly $r > 0$ nonzero coefficients and that $\pi : \{1,\dots,p\} \longrightarrow \{1,\dots,p\}$ is permutation which gives the order of magnitudes for Dy, i.e. $d_{\pi(1)}|y|_{\pi(1)} \ge \dots \ge d_{\pi(p)}|y|_{\pi(p)}$. Basing on our previous observations, we get an important

Corollary E.4. *If b^* is the solution to (E.1) having exactly $r > 0$ nonzero coefficients and π is permutation which places components of $D|y|$ in a nonincreasing order, i.e. $d_{\pi(i)}|y|_{\pi(i)} = |Dy|_{(i)}$ for $i = 1,\dots,p$, then the support of b^* is composed of the set $\{\pi(1),\dots,\pi(r)\}$.*

The next three lemmas were proven in (Bogdan et al., 2013) in situation when $d_1 = \dots = d_p = 1$. We will follow the reasoning from this paper to prove the generalized claims. The main difference is that in general case the solution to the considered problem (E.1) does not have to be nonincreasingly ordered, under assumption that $d_1 y_1 \ge \dots \ge d_p y_p \ge 0$ (which is the case for $d_1 = \dots = d_p = 1$). This means that generalizations of proofs presented in (Bogdan et al., 2013) are not straightforward.

Lemma E.5. *Consider nonnegative sequence $\{y_i\}_{i=1}^p$ and sequence of positive numbers $\{d_i\}_{i=1}^p$ such that $d_1 y_1 \geq \ldots \geq d_p y_p$. If b^* is solution to problem (E.1) having exactly r nonzero entries, then for every $j \leq r$ it holds that*

$$\sum_{i=j}^{r}(d_i y_i - \lambda_i) > 0 \tag{E.5}$$

and for every $j \geq r+1$

$$\sum_{i=r+1}^{j}(d_i y_i - \lambda_i) \leq 0. \tag{E.6}$$

Proof. From Proposition E.3 we know that $b_i^* > 0$ for $i \in \{1, \ldots, r\}$. Let us define

$$\widetilde{b}_i := \begin{cases} b_i^* - h, i \in \{j, \ldots, r\} \\ b_i^*, \text{otherwise.} \end{cases},$$

where we restrict only to sufficiently small values of h, so as to the condition $\widetilde{b}_i > 0$ is met for all i from $\{j, \ldots, r\}$. For such h we have $b_{(r+1)}^* = \ldots = b_{(p)}^* = \widetilde{b}_{(r+1)} = \ldots = \widetilde{b}_{(p)} = 0$. Therefore there exists permutation $\pi : \{1, \ldots, r\} \longrightarrow \{1, \ldots, r\}$ such as $\sum_{i=1}^{r} \lambda_i \widetilde{b}_{(i)} = \sum_{i=1}^{r} \lambda_{\pi(i)} \widetilde{b}_i$. For such permutation we have

$$\begin{aligned} J_\lambda(b^*) - J_\lambda(\widetilde{b}) = & \sum_{i=1}^{r} \lambda_i b_{(i)}^* - \sum_{i=1}^{r} \lambda_i \widetilde{b}_{(i)} = \sum_{i=1}^{r} \lambda_i b_{(i)}^* - \sum_{i=1}^{r} \lambda_{\pi(i)} \widetilde{b}_i \\ \geq & \sum_{i=1}^{r} \lambda_{\pi(i)} b_i^* - \sum_{i=1}^{r} \lambda_{\pi(i)} \widetilde{b}_i = h \sum_{i=j}^{r} \lambda_{\pi(i)} \geq h \sum_{i=j}^{r} \lambda_i, \end{aligned} \tag{E.7}$$

where the first inequality follows from the rearrangement inequality and second is the consequence of monotonicity of $\{\lambda_i\}_{i=1}^p$. We also have

$$\begin{aligned} \|y - Db^*\|_2^2 - \|y - D\widetilde{b}\|_2^2 = & \sum_{i=j}^{r}(y_i - d_i b_i^*)^2 - \sum_{i=j}^{r}(y_i - d_i b_i^* + d_i h)^2 \\ & = 2h \sum_{i=j}^{r}(d_i^2 b_i^* - d_i y_i) - h^2 \sum_{i=j}^{r} d_i^2. \end{aligned} \tag{E.8}$$

Optimality of b^*, (E.7) and (E.8) yield

$$0 \geq f(b^*) - f(\widetilde{b}) \geq h \sum_{i=j}^{r}(d_i^2 b_i^* - d_i y_i + \lambda_i) - \frac{1}{2} h^2 \sum_{i=j}^{r} d_i^2, \tag{E.9}$$

for each h from the interval $[0, \varepsilon]$, where $\varepsilon > 0$ is some (sufficiently small) value. This gives $\sum_{i=j}^{r}(d_i^2 b_i^* - d_i y_i + \lambda_i) \leq 0$ and consequently

$$\sum_{i=j}^{r}(d_i y_i - \lambda_i) \geq \sum_{i=j}^{r} d_i^2 b_i^* > 0. \tag{E.10}$$

To prove claim (E.6), consider a new sequence defined as $\widetilde{b}_i := \begin{cases} h, i \in \{r+1, \ldots, j\} \\ b_i^*, \text{otherwise.} \end{cases}$.
We will restrict our attention only to $0 < h < \min\{b_i^* : \ i \leq r\}$, so as to $b_{(\cdot)}^*$ and $\widetilde{b}_{(\cdot)}$ are given by applying the same permutation to b^* and $\widetilde{b}$, respectively. Moreover, for each i from $\{r+1, \ldots, j\}$ it holds $\widetilde{b}_{(i)} = \widetilde{b}_i = h$. From optimality of b^*

$$0 \geq f(b^*) - f(\widetilde{b}) = \frac{1}{2}\sum_{i=r+1}^{j}\left(y_i^2 - (y_i - d_i h)^2\right) - \sum_{i=r+1}^{j}\lambda_i h = h\sum_{i=r+1}^{j}(d_i y_i - \lambda_i) - \frac{1}{2}h^2\sum_{i=r+1}^{j} d_i^2,$$

for all considered h, which leads to (E.6). ∎

Lemma E.6. *Let b^* be solution to problem (E.1) with nonnegative, nonincreasing sequence $\{\lambda_i\}_{i=1}^p$. Let $R(b^*)$ be number of all nonzeros in b^* and $r \geq 1$. Then, for any $i \in \{1, \ldots, p\}$*

$$\left\{y : \ b_i^* \neq 0 \ \textit{and} \ R(b^*) = r\right\} = \left\{y : \ d_i|y_i| > \lambda_r \ \textit{and} \ R(b^*) = r\right\}. \tag{E.11}$$

Proof. Suppose that b^* has $r > 0$ nonzero coefficients and let π be permutation which places components of $D|y|$ in a non-increasing order. From Corollary E.4 it holds that $\{i : \ b_i^* \neq 0\} = \{\pi(1), \ldots, \pi(r)\}$. Define $\widetilde{y} := P_\pi S y$ and $\widetilde{D} := P_\pi D P_\pi^\intercal$, for S being the diagonal matrix such as $S_{i,i} = sgn(y_i)$. Then $P_\pi S b^*$ is solution to problem

$$\underset{b}{\operatorname{argmin}}\left\{ \ \frac{1}{2}\left\|\widetilde{y} - \widetilde{D}b\right\|_2^2 \ + \ J_\lambda(b) \ \right\}, \tag{E.12}$$

which satisfies the assumptions of Lemma E.5. Taking $j = r$ in (E.5) and $j = r+1$ in $(E.6)$ we immediately get

$$d_{\pi(r)}|y|_{\pi(r)} > \lambda_r \ \text{ and } \ d_{\pi(r+1)}|y|_{\pi(r+1)} \leq \lambda_{r+1}. \tag{E.13}$$

We will now show that $\left\{y : \ b_i^* \neq 0 \text{ and } R(b^*) = r\right\} \subset \left\{y : \ d_i|y_i| > \lambda_r \text{ and } R(b^*) = r\right\}$.

Fix $i \in \{1, \ldots, p\}$ and suppose that b_i^* is nonzero coefficient. Then $i \in \{\pi(1), \ldots, \pi(r)\}$ and therefore $d_i|y_i| \geq d_{\pi(r)}|y|_{\pi(r)} > \lambda_r$, thanks to first inequality from (E.13). To show the second inclusion assume that $d_i|y_i| > \lambda_r$. Then, from the second inequality in (E.13), $d_i|y_i| > \lambda_{r+1} \geq d_{\pi(r+1)}|y|_{\pi(r+1)}$, which gives $i \in \{\pi(1), \ldots, \pi(r)\}$. ■

Lemma E.7. *For given sequence $\{y_i\}_{i=1}^p$, sequence of positive numbers $\{d_i\}_{i=1}^p$, nonincreasing, nonnegative sequence $\{\lambda_i\}_{i=1}^p$ and fixed $j \in \{1, \ldots, p\}$, consider the following procedure*

- ***P.1*** *define $\widetilde{y} := (y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_p)^\mathsf{T}$, $\widetilde{D} := diag(d_1, \ldots, d_{j-1}, d_{j+1}, \ldots, d_p)$, $\widetilde{d}_i := \widetilde{D}_{i,i}$ for $i = 1, \ldots, p-1$ and $\widetilde{\lambda} := (\lambda_2, \ldots, \lambda_p)^\mathsf{T}$;*
- ***P.2*** *find $\widetilde{b}^* := \operatorname{argmin}_{b \in \mathbb{R}^{p-1}} \frac{1}{2}\left\|\widetilde{y} - \widetilde{D}b\right\|_2^2 + J_{\widetilde{\lambda}}(b)$;*
- ***P.3*** *define $\widetilde{R}^j(\widetilde{b}^*) := |\{i : \ \widetilde{b}_i^* \neq 0\}|$.*

Then for $r \geq 1$ it holds $\big\{y : \ d_j|y_j| > \lambda_r$ and $R(b^*) = r\big\} \subset \big\{y : \ d_j|y_j| > \lambda_r$ and $\widetilde{R}^j(\widetilde{b}^*) = r - 1\big\}$.

Proof. We have to show that solution $\widetilde{b}^*$ to problem

$$\underset{b}{\text{minimize}} \;\; F(b) := \frac{1}{2}\sum_{i=1}^{p-1}\Big(\widetilde{y}_i - \widetilde{d}_i b_i\Big)^2 + \sum_{i=1}^{p-1}\widetilde{\lambda}_i b_{(i)} \tag{E.14}$$

has exactly $r-1$ nonzero coefficients. From Proposition E.1 we know that the change of signs of y_i's does not affect the support, hence without loss of generality we can assume that $\widetilde{y} \succeq 0$, and $\widetilde{b}^* \succeq 0$ as a result (from Proposition E.2). We will start with situation when $d_1 y_1 \geq \ldots \geq d_p y_p$ and consequently $\widetilde{d}_1\widetilde{y}_1 \geq \ldots \geq \widetilde{d}_{p-1}\widetilde{y}_{p-1}$. If j is fixed index such as $d_j|y_j| > \lambda_r$ and $R(b^*) = r$, this gives

$$j \in \{1, \ldots, r\}. \tag{E.15}$$

To show that solution to (E.14) has at least $r-1$ nonzero entries, suppose by contradiction that $\widetilde{b}^*$ has exactly $k-1$ nonzero entries with $k < r$. Let us define $\widehat{b} \in \mathbb{R}^{p-1}$ as

$$\widehat{b}_i := \begin{cases} h, \ i \in \{k, \dots, r-1\} \\ \widetilde{b}_i^*, \text{otherwise} \end{cases}, \tag{E.16}$$

where $0 < h < \min\{\widetilde{b}_1^*, \dots, \widetilde{b}_{k-1}^*\}$. Then

$$F(\widetilde{b}^*) - F(\widehat{b}) = h \sum_{i=k}^{r-1} (\widetilde{d}_i \widetilde{y}_i - \widetilde{\lambda}_i) - h^2 \sum_{i=k}^{r-1} \frac{1}{2} \widetilde{d}_i^2. \tag{E.17}$$

Now

$$\sum_{i=k}^{r-1} (\widetilde{d}_i \widetilde{y}_i - \widetilde{\lambda}_i) = \sum_{i=k+1}^{r} (\widetilde{d}_{i-1} \widetilde{y}_{i-1} - \lambda_i) \geq \sum_{i=k+1}^{r} (d_i y_i - \lambda_i) > 0, \tag{E.18}$$

where the first equality follows from $\widetilde{\lambda}_i = \lambda_{i+1}$, the first inequality from $\widetilde{d}_{i-1} \widetilde{y}_{i-1} \geq d_i y_i$ and the second from Lemma E.5. If h is small enough, we get $F(\widehat{b}) < F(\widetilde{b}^*)$ which leads to contradiction.

Suppose now by contradiction that $\widetilde{b}^*$ has k nonzero entries with $k \geq r$ and define

$$\widehat{b}_i := \begin{cases} \widetilde{b}_i^* - h, i \in \{r, \dots, k\} \\ \widetilde{b}_i^*, \qquad \text{otherwise} \end{cases}. \tag{E.19}$$

Analogously to (E.7), we get $J_{\widetilde{\lambda}}(\widetilde{b}^*) - J_{\widetilde{\lambda}}(\widehat{b}) \geq h \sum_{i=r}^{k} \widetilde{\lambda}_i$ and consequently

$$F(\widetilde{b}^*) - F(\widehat{b}) \geq h \left[\sum_{i=r}^{k} (\widetilde{\lambda}_i - \widetilde{d}_i \widetilde{y}_i) + \sum_{i=r}^{k} \widetilde{d}_i^2 \widetilde{b}_i^* \right] - \frac{1}{2} h^2 \sum_{i=r}^{k} \widetilde{d}_i^2. \tag{E.20}$$

Now

$$\sum_{i=r}^{k} (\widetilde{\lambda}_i - \widetilde{d}_i \widetilde{y}_i) = \sum_{i=r+1}^{k+1} (\lambda_i - d_i y_i) \geq 0, \tag{E.21}$$

where the first equality follows from definition of $\widetilde{\lambda}$ and $(E.15)$, while the inequality follows from Lemma E.5. If h is small enough, we get $F(\widehat{b}) < F(\widetilde{b}^*)$, which contradicts the optimality of $\widetilde{b}^*$.

Consider now general situation, i.e. without assumption concerning the order of $D|y|$. Suppose that π, with corresponding matrix P_π, is permutation which orders $D|y|$. Define $y_\pi := P_\pi y$ and $D_\pi := P_\pi D P_\pi^\intercal$. Applying the procedure described in the statement of Lemma simultaneously to (y, D, λ) for j, and to (y_π, D_π, λ) for $\pi(j)$ we end with $\big(\widetilde{y}, \widetilde{D}, \widetilde{\lambda}, \widetilde{R}_1^j(\widetilde{b}^*)\big)$ and $\big(\widetilde{y_\pi}, \widetilde{D_\pi}, \widetilde{\lambda}, \widetilde{R}_2^{\pi(j)}(\widetilde{b}_\pi^*)\big)$. It is straightforward to see, that there exists permutation $\widetilde{\pi} : \{1, \dots, p-1\} \longrightarrow \{1, \dots, p-1\}$ such that $\widetilde{y_\pi} = P_{\widetilde{\pi}} \widetilde{y}$ and

$\widetilde{D_\pi} = P_{\widetilde{\pi}} \widetilde{D} P_{\widetilde{\pi}}^{\mathsf{T}}$. From Proposition E.1 we have that $\widetilde{b}_\pi^* = P_{\widetilde{\pi}} \widetilde{b}^*$ and $\widetilde{R}_1^j(\widetilde{b}^*) = \widetilde{R}_2^{\pi(j)}(\widetilde{b}_\pi^*)$. Moreover, from the first part of proof $\widetilde{R}_2^{\pi(j)}(\widetilde{b}_\pi^*) = r - 1$, which gives the claim. ■

F Minimax properties of gSLOPE

Proof of Theorem 2.6. Once again we will employ the equivalent formulation of gSLOPE under assumption about orthogonality at the groups level, i.e. problem (2.12), and we will consider statistically equivalent model $\widetilde{y} \sim \mathcal{N}\big(\widetilde{\beta},\ \sigma^2 \mathbf{I}_{\widetilde{p}}\big)$, with $\widetilde{\beta}_{\mathbb{I}_i} = R_i \beta_{I_i}$, $i = 1, \ldots, m$. Then $[\![\beta]\!]_{I,X} = [\![\widetilde{\beta}]\!]_{\mathbb{I}}$ and for solution b^* to (2.12) it holds $[\![b^*]\!]_{\mathbb{I}} = [\![\beta^{\mathrm{gS}}]\!]_{I,X}$ for any solution β^{gS} to problem (2.2). Without loss of generality, assume $\sigma = 1$. Note that $\|\widetilde{y}_{\mathbb{I}_i}\|_2^2$ is distributed as the noncentral $\chi_{l_i}^2(\|\widetilde{\beta}_{\mathbb{I}_i}\|_2^2)$, where $\|\widetilde{\beta}_{\mathbb{I}_i}\|_2^2$ is the noncentrality.

The lower bound of the minimax risk can be obtained as follows. For each $\mathbb{I}_i$, only $\widetilde{\beta}_j$ with the smallest index $j \in \mathbb{I}_i$ is *possibly* nonzero and the rest $l_i - 1$ components of $\widetilde{\beta}_{\mathbb{I}_i}$ are fixed to be zero. Then, this is reduced to a simple Gaussian sequence model with length m and sparsity at most k. Given the condition $k/m \to 0$, this classical sequence model has minimax risk $(1 + o(1))2k\log(m/k)$ (see e.g. Donoho and Johnstone, 1994).

Our next step is to evaluate the worst risk of gSLOPE over the nearly black object. We would completes the proof if we show this worst risk is bounded above by $(1 + o(1))2k\log(m/k)$. For simplicity, assume that $\|\widetilde{\beta}_{\mathbb{I}_i}\|_2 = 0$ for all $i \geq k+1$ and write $\mu_i = \|\widetilde{\beta}_{\mathbb{I}_i}\|_2, \zeta_i = \|\widetilde{y}_{\mathbb{I}_i}\|_2 \sim \chi_{l_i}(\mu_i^2)$. Denote by $\widehat{\zeta}$ the SLOPE solution. Then, the risk is

$$\mathbb{E}\|\widehat{\zeta} - \mu\|_2^2 = \mathbb{E}\sum_{i=1}^{k}(\widehat{\zeta}_i - \mu_i)_2^2 + \mathbb{E}\sum_{i=k+1}^{m}\widehat{\zeta}_i^2.$$

Then, it suffices to show

$$\mathbb{E}\left[\sum_{i=1}^{k}(\widehat{\zeta}_i - \mu_i)^2\right] \leq (1 + o(1))2k\log(m/k) \tag{F.1}$$

and

$$\mathbb{E}\left[\sum_{i=k+1}^{m}\widehat{\zeta}_i^2\right] = o(1)2k\log(m/k). \tag{F.2}$$

Below, Lemmas F.1, F.2, and F.3 together give (F.2). The remaining part of this proof serves to validate (F.1). To start with, we employ the representation $\zeta_i^2 = (\xi_{i1} + \mu_i)^2 + \xi_{i2}^2 + \cdots + \xi_{il_i}^2$ for i.i.d. $\xi_{ij} \sim \mathcal{N}(0,1)$ (we can assume this representation without loss of generality, since the distribution of $(\xi_{i1}+a_1)^2+(\xi_{i2}+a_2)^2+\cdots+(\xi_{il_i}+a_{l_i})^2$ depends only on the non-centrality $a_1^2+\cdots+a_{l_i}^2$). As in the proof of Lemma 3.2 in (Su and Candès, 2016), we get

$$\begin{aligned}\sum_{i=1}^{k}(\widehat{\zeta}_i - \mu_i)^2 &\le \left(\|\widehat{\zeta}_{[1:k]} - \zeta_{[1:k]}\|_2 + \|\zeta_{[1:k]} - \mu_{[1:k]}\|_2\right)^2 \\ &\le \left(\|\lambda_{[1:k]}\|_2 + \|\zeta_{[1:k]} - \mu_{[1:k]}\|_2\right)^2.\end{aligned} \tag{F.3}$$

As l is fixed and $k/m \to 0$, (Inglot, 2010) gives $\lambda_i \sim \sqrt{2\log\frac{m}{qi}}$ for all $i \le k$. From this we know

$$\|\lambda_{[1:k]}\|_2^2 = \sum_{i=1}^{k}\lambda_i^2 \sim 2k\log\frac{m}{k}. \tag{F.4}$$

Next, we see

$$\begin{aligned}\left|\sqrt{(\xi_{i1}+\mu_i)^2 + \xi_{i2}^2 + \cdots + \xi_{il_i}^2} - \mu_i\right| &\le \sqrt{\xi_{i2}^2 + \cdots + \xi_{il_i}^2} + |\xi_{i1}| \\ &\le 2\sqrt{\xi_{i1}^2 + \xi_{i2}^2 + \cdots + \xi_{il_i}^2} \equiv 2\|\xi_i\|_2,\end{aligned}$$

which yields

$$\|\zeta_{[1:k]} - \mu_{[1:k]}\|_2^2 \le 4\sum_{i=1}^{k}\|\xi_i\|_2^2 \tag{F.5}$$

Note that $\sum_{i=1}^{k}\|\xi_i\|_2^2$ is distributed as the chi-square with $l_1 + \cdots + l_k \le lk$ degrees of freedom. Taking (F.4) and (F.5) together, from (F.3) we get

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{k}(\widehat{\zeta}_i - \mu_i)^2\right] &\le \|\lambda_{[1:k]}\|_2^2 + \mathbb{E}\|\zeta_{[1:k]} - \mu_{[1:k]}\|_2^2 + 2\|\lambda_{[1:k]}\|_2\mathbb{E}\|\zeta_{[1:k]} - \mu_{[1:k]}\|_2 \\ &\le (1+o(1))2k\log\frac{m}{k} + 4lk + 2\sqrt{(1+o(1))2k\log\frac{m}{k}}\cdot\sqrt{4lk} \\ &\sim (1+o(1))2k\log\frac{m}{k},\end{aligned}$$

where the last step makes use of $m/k \to \infty$. This establishes (F.1) and consequently completes the proof.

$\blacksquare$

The following three lemmas aim to prove (F.2). Denote by $\zeta_{(1)} \geq \cdots \geq \zeta_{(m-k)}$ the order statistics of $\zeta_{k+1}, \ldots, \zeta_m$. Recall that $\zeta_i \sim \chi_{l_i}$ for $i \geq k+1$. As in the proof of Lemma 3.3 in (Su and Candès, 2016), we have

$$\sum_{i=k+1}^{m} \widehat{\zeta}_i^2 \leq \sum_{i=1}^{m-k} (\zeta_{(i)} - \lambda_{k+i})_+^2,$$

where $x_+ = \max\{x, 0\}$. For a sufficiently large constant $A > 0$ and sufficiently small constant $\alpha > 0$ both to be specified later, we partition the sum into three parts:

$$\sum_{i=1}^{m-k} \mathbb{E}(\zeta_{(i)} - \lambda_{k+i})_+^2 = \sum_{i=1}^{\lfloor Ak \rfloor} \mathbb{E}(\zeta_{(i)} - \lambda_{k+i})_+^2 + \sum_{i=\lceil Ak \rceil}^{\lfloor \alpha m \rfloor} \mathbb{E}(\zeta_{(i)} - \lambda_{k+i})_+^2 + \sum_{i=\lceil \alpha m \rceil}^{m-k} \mathbb{E}(\zeta_{(i)} - \lambda_{k+i})_+^2$$

The three lemmas, respectively, show that each part is negligible compared with $2k \log(m/k)$. We indeed prove a stronger version in which the order statistics $\zeta_{(1)} \geq \cdots \geq \zeta_{(m-k)} \geq \zeta_{(m-k+1)} \geq \cdots \geq \zeta_{(m)}$ come from m i.i.d. χ_l. Let $U_1, \ldots, U_m$ be i.i.d. uniform random variables on $(0, 1)$, and $U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(m)}$ be the *increasing* order statistics. Hence, we have the representation $\zeta_{(i)} = F_{\chi_l}^{-1}(1 - U_{(i)})$. Below in the proof of Lemma F.1, we write $a_m \lesssim b_m$ for two positive sequences a_m and b_m if there exists some positive constant C such that $a_m \leq C b_m$ for all m.

Lemma F.1. *Under the preceding conditions, for any $A > 0$ we have*

$$\frac{1}{2k \log(m/k)} \sum_{i=1}^{\lfloor Ak \rfloor} \mathbb{E}(\zeta_{(i)} - \lambda_{k+i})_+^2 \to 0.$$

Proof of Lemma F.1. Recognizing that l is fixed, from (Inglot, 2010) it follows that

$$\lim_{q_1, q_2 \to 0, q_2 > q_1} \frac{F_{\chi_l}^{-1}(1 - q_1) - F_{\chi_l}^{-1}(1 - q_2)}{\sqrt{2 \log \frac{1}{q_1}} - \sqrt{2 \log \frac{1}{q_2}}} = 1.$$

We also know that ζ_i is distributed as $F_{\chi_l}^{-1}(1 - U_{(i)})$. Making use of these facts, we get

$$
\begin{aligned}
\mathbb{E}(\zeta_{(i)} - \lambda_{k+i})_+^2 &= \mathbb{E}(F_{\chi_l}^{-1}(1 - U_{(i)}) - F_{\chi_l}^{-1}(1 - q(k+i)/m))_+^2 \\
&\sim \mathbb{E}\left(\sqrt{2\log\frac{1}{U_{(i)}}} - \sqrt{2\log\frac{m}{q(k+i)}}\right)_+^2 \\
&\le \mathbb{E}\left(\sqrt{2\log\frac{1}{U_{(i)}}} - \sqrt{2\log\frac{m}{q(k+i)}}\right)^2 \\
&\lesssim \mathbb{E}\left(\frac{\log^2(q(k+i)/mU_{(i)})}{\log(m/q(k+i))}\right).
\end{aligned}
$$

Now, we proceed to evaluate

$$
\mathbb{E}\left[\log^2\frac{q(k+i)}{mU_{(i)}}\right] = \log^2\frac{q(k+i)}{m} + \mathbb{E}\log^2 U_{(i)} - 2\log\frac{q(k+i)}{m}\mathbb{E}\log U_{(i)}.
$$

Here, $U_{(i)}$ follows the Beta$(i, m+1-i)$ distribution and, hence, it has mean $i/(m+1)$. Similar results concerning the logarithm of $U_{(i)}$ are given in (Abramowitz and Stegun, 1964):

$$
\begin{aligned}
\mathbb{E}\log U_{(i)} &= -\log\frac{m+1}{i} + \delta_1, \\
\mathbb{E}\log^2 U_{(i)} &= \left(\log\frac{m+1}{i} - \delta_1\right)^2 + \frac{1}{i} - \frac{1}{m+1} + \delta_2
\end{aligned}
$$

for some $\delta_1 = O(1/i)$ and $\delta_2 = O(1/i^2)$. Thus we can evaluate $\mathbb{E}\log^2\frac{q(k+i)}{mU_{(i)}}$ as

$$
\begin{aligned}
\mathbb{E}\log^2\frac{q(k+i)}{mU_{(i)}} &= \log^2\frac{q(k+i)}{m} - 2\log\frac{q(k+i)}{m}\mathbb{E}\log U_{(i)} + \mathbb{E}\log^2 U_{(i)} \\
&= \log^2\frac{q(k+i)}{m} + 2\log\frac{q(k+i)}{m}\left(\log\frac{m+1}{i} - \delta_1\right) + \left(\log\frac{m+1}{i} - \delta_1\right)^2 + \frac{1}{i} - \frac{1}{m+1} + \delta_2 \\
&= \log^2\frac{q(k+i)(m+1)}{im} - 2\delta_1\log\frac{q(k+i)(m+1)}{im} + \frac{1}{i} - \frac{1}{m+1} + \delta_1^2 + \delta_2.
\end{aligned}
$$

Hence, we get

$$
\begin{aligned}
&\sum_{i=1}^{\lfloor Ak\rfloor}\mathbb{E}(\zeta_{(i)}-\lambda_{k+i})_+^2\\
&\lesssim \frac{1}{\log\frac{m}{q(A+1)k}}\left(\underbrace{\sum_{i=1}^{\lfloor Ak\rfloor}\log^2\frac{q(k+i)(m+1)}{im}}_{\text{I}}-\underbrace{\sum_{i=1}^{\lfloor Ak\rfloor}2\delta_1\log\frac{q(k+i)(m+1)}{im}}_{\text{II}}+\underbrace{\sum_{i=1}^{\lfloor Ak\rfloor}(\frac{1}{i}-\frac{1}{m+1}+\delta_1^2+\delta_2)}_{\text{III}}\right)\\
&\le \frac{1}{\log\frac{m}{q(A+1)k}}(\text{I}+|\text{II}|+|\text{III}|).
\end{aligned}
$$

Since $\frac{m}{q(A+1)k}\to\infty$, the proof would be completed once we show $\text{I}, |\text{II}|$, and $|\text{III}|$ are bounded. To this end, first note that

$$
\begin{aligned}
\text{I}&=\sum_{i=1}^{\lfloor Ak\rfloor}\log^2\frac{q(k+i)(m+1)}{im}\\
&\le\sum_{i=1}^{\lfloor Ak\rfloor}\max\left\{k\int_{(i-1)/k}^{i/k}\log^2\frac{q(m+1)(1+x)}{mx}\mathrm{d}x, k\int_{i/k}^{(i+1)/k}\log^2\frac{q(m+1)(1+x)}{mx}\mathrm{d}x\right\}\\
&\le 2k\int_0^{A+1}\log^2\frac{q(m+1)(1+x)}{mx}\mathrm{d}x\asymp k=o\left(2k\log\frac{m}{k}\right).
\end{aligned}
$$

The second term II obeys

$$
\begin{aligned}
|\text{II}|&\le\sum_{i=1}^{\lfloor Ak\rfloor}2\Big|\delta_1\log\frac{q(k+i)(m+1)}{im}\Big|\lesssim\sum_{i=1}^{\lfloor Ak\rfloor}\frac{1}{i}\Big|\log\frac{q(k+i)(m+1)}{im}\Big|\\
&\le\sum_{i=1}^{\lfloor Ak\rfloor}\max\left\{k\int_{(i-1)/k}^{i/k}\Big|\log\frac{q(m+1)(1+x)}{mx}\Big|\mathrm{d}x, k\int_{i/k}^{(i+1)/k}\Big|\log\frac{q(m+1)(1+x)}{mx}\Big|\mathrm{d}x\right\}\\
&\le 2k\int_0^{A+1}\Big|\log\frac{q(m+1)(1+x)}{mx}\Big|\mathrm{d}x\asymp k=o\left(2k\log\frac{m}{k}\right),
\end{aligned}
$$

where we use the fact that $\int_0^{A+1}\Big|\log\frac{q(m+1)(1+x)}{mx}\Big|\mathrm{d}x$ is bounded by some constant. The last term is simply bounded as

$$
\begin{aligned}
|\mathrm{III}| &\le \sum_{i=1}^{\lfloor Ak \rfloor} \Big| \frac{1}{i} - \frac{1}{m+1} + \delta_1^2 + \delta_2 \Big| \\
&= \sum_{i=1}^{\lfloor Ak \rfloor} \Big| \frac{1}{i} - \frac{1}{m+1} + O(1/i^2) + O(1/i^2) \Big| \\
&\lesssim \sum_{i=1}^{\lfloor Ak \rfloor} \frac{1}{i} \;\lesssim\; \log(Ak) \;=\; o\left(2k \log \frac{m}{k}\right),
\end{aligned}
$$

where the last equality is due to the fact that $m/k \to \infty$. Combining these established bounds on I, II, and III finishes proof. ∎

Lemma F.2. *Under the preceding conditions, let A be any constant satisfying $q(1+A)/A < 1$ and α be sufficiently small such that $l/\lambda_{k+\lfloor \alpha m \rfloor} < 1/2$. Then,*

$$
\frac{1}{2k\log(m/k)} \sum_{i=\lceil Ak \rceil}^{\lfloor \alpha m \rfloor} \mathbb{E}(\zeta_{(i)} - \lambda_{k+i})_+^2 \to 0.
$$

Proof of Lemma F.2. Note that $\lambda_{k+\lfloor \alpha m \rfloor} \sim \sqrt{2\log \frac{m}{q(k+\lfloor \alpha m \rfloor)}} \sim \sqrt{2\log \frac{1}{q\alpha}}$. So it is clear that such α exists. Pick any fixed i between $\lceil Ak \rceil$ and $\lfloor \alpha m \rfloor$. As in the proof of Lemma A.4 in (Su and Candès, 2016), denote by $\alpha_u = \mathbb{P}(\chi_l > \lambda_{k+i} + u)$. Note that

$$
\begin{aligned}
\alpha_u &= \mathbb{P}(\chi_l > \lambda_{k+i} + u) = \int_{(\lambda_{k+i}+u)^2}^{\infty} \frac{1}{\mathrm{e}^{l/2}\Gamma(l/2)} x^{l/2-1} \mathrm{e}^{-x/2} \mathrm{d}x \\
&= \int_{\lambda_{k+i}^2}^{\infty} \frac{1}{\mathrm{e}^{l/2}\Gamma(l/2)} \left(\frac{(\lambda_{k+i}+u)^2}{\lambda_{k+i}^2} y \right)^{l/2-1} \exp\left(-\frac{(\lambda_{k+i}+u)^2}{2\lambda_{k+i}^2} y \right) \mathrm{d} \frac{(\lambda_{k+i}+u)^2}{\lambda_{k+i}^2} y \\
&= \left(1 + \frac{u}{\lambda_{k+i}}\right)^l \int_{\lambda_{k+i}^2}^{\infty} \frac{1}{\mathrm{e}^{l/2}\Gamma(l/2)} y^{l/2-1} \exp\left(-\frac{(\lambda_{k+i}+u)^2}{2\lambda_{k+i}^2} y \right) \mathrm{d}y \\
&\le \left(1 + \frac{u}{\lambda_{k+i}}\right)^l \mathrm{e}^{-\lambda_{k+i} u} \int_{\lambda_{k+i}^2}^{\infty} \frac{1}{e^{l/2}\Gamma(l/2)} y^{l/2-1} \mathrm{e}^{-y/2} \mathrm{d}y \\
&= \left(1 + \frac{u}{\lambda_{k+i}}\right)^l \mathrm{e}^{-\lambda_{k+i} u} \alpha_0 \\
&\le \exp\left(\frac{l}{\lambda_{k+i}} u - \lambda_{k+i} u \right) \alpha_0.
\end{aligned}
$$

With the proviso that $l/\lambda_{k+\lfloor \alpha m \rfloor} < 1/2 < \lambda_{k+\lfloor \alpha m \rfloor}/2$, it follows that

$$\alpha_u \le \mathrm{e}^{-\lambda_{k+i} u/2} \alpha_0.$$

The remaining proof follows from exactly the same reasoning as that of Lemma A.4 in (Su and Candès, 2016).

■

Lemma F.3. *Under the preceding conditions, for any constant $\alpha > 0$ we have*

$$\frac{1}{2k\log(m/k)} \sum_{i=\lceil \alpha m \rceil}^{m-k} \mathbb{E}(\zeta_{(i)} - \lambda_{k+i})_+^2 \to 0.$$

Proof of Lemma F.3. Recognizing that the value of the summation increases as α decreases, we only prove the lemma for sufficiently small α. In the case of $U_{(i)} \ge \alpha/3$, we get

$$\begin{aligned}
(\zeta_{(i)} - \lambda_{k+i})_+ &= \left(F_{\chi_l}^{-1}(1 - U_{(i)}) - F_{\chi_l}^{-1}(1 - q(k+i)/m)\right)_+ \\
&\asymp \left(1 - U_{(i)} - (1 - q(k+i)/m)\right)_+ \\
&= (q(k+i)/m - U_{(i)})_+,
\end{aligned}$$

since both $U_{(i)}$ and $q(k+i)/m$ are bounded below away from zero. Otherwise, we use the trivial inequality $(\zeta_{(i)} - \lambda_{k+i})_+ \le \zeta_{(i)}$. In either case, we get

$$\begin{aligned}
(\zeta_{(i)} - \lambda_{k+i})_+^2 &\lesssim \zeta_{(i)}^2 \mathbf{1}_{U_{(i)} < \frac{\alpha}{3}} + \left(\frac{q(k+i)}{m} - U_{(i)}\right)_+^2 \\
&= \left(F_{\chi_l}^{-1}(1 - U_{(i)})\right)^2 \mathbf{1}_{U_{(i)} < \frac{\alpha}{3}} + \left(\frac{q(k+i)}{m} - U_{(i)}\right)_+^2 \\
&\asymp 2\log\left(\frac{1}{U_{(i)}}\right) \mathbf{1}_{U_{(i)} < \frac{\alpha}{3}} + \left(\frac{q(k+i)}{m} - U_{(i)}\right)_+^2 \\
&\lesssim \log\left(\frac{1}{U_{(i)}}\right) \mathbf{1}_{U_{(i)} < \frac{\alpha}{3}} + \mathbf{1}_{U_{(i)} \le \frac{q(k+i)}{m}}.
\end{aligned}$$

Hence,

$$\sum_{i=\lceil \alpha m \rceil}^{m-k} \mathbb{E}(\zeta_{(i)} - \lambda_{k+i})_+^2 \lesssim \sum_{i=\lceil \alpha m \rceil}^{m-k} \mathbb{E}\left(\log\left(\frac{1}{U_{(i)}}\right); U_{(i)} < \frac{\alpha}{3}\right) + \sum_{i=\lceil \alpha m \rceil}^{m-k} \mathbb{P}\left(U_{(i)} \le \frac{q(k+i)}{m}\right)$$

In the remaining proof we aim to show

$$\sum_{i=\lceil \alpha m \rceil}^{m-k} \mathbb{E}\left(\log\left(\frac{1}{U_{(i)}}\right); U_{(i)} < \frac{\alpha}{3}\right) \to 0 \tag{F.6}$$

and
$$\sum_{i=\lceil \alpha m \rceil}^{m-k} \mathbb{P}\left(U_{(i)} \le \frac{q(k+i)}{m} \right) \to 0. \tag{F.7}$$

This is more than we need since $2k\log(m/k) \to \infty$.

Each summand of (F.6) is bounded above by

$$\begin{aligned}
\mathbb{E}\left(\log\left(\frac{1}{U_{(\lceil \alpha m \rceil)}} \right); U_{(\lceil \alpha m \rceil)} < \frac{\alpha}{3} \right) &= \int_0^{\frac{\alpha}{3}} \frac{x^{\lceil \alpha m \rceil - 1}(1-x)^{m-\lceil \alpha m \rceil} \log \frac{1}{x}}{\mathrm{B}(\lceil \alpha m \rceil, m+1-\lceil \alpha m \rceil)} \mathrm{d}x \\
&\le \int_0^{\frac{\alpha}{3}} \frac{x^{\lceil \alpha m \rceil - 1} \log \frac{1}{x}}{\mathrm{B}(\lceil \alpha m \rceil, m+1-\lceil \alpha m \rceil)} \mathrm{d}x \\
&= \frac{1}{\lceil \alpha m \rceil^2 \, \mathrm{B}(\lceil \alpha m \rceil, m+1-\lceil \alpha m \rceil)} \int_0^{(\frac{\alpha}{3})^{\lceil \alpha m \rceil}} \log \frac{1}{y} \, \mathrm{d}y \\
&\sim \frac{(\alpha/3)^{\lceil \alpha m \rceil} \log \frac{3}{\alpha}}{\lceil \alpha m \rceil \, \mathrm{B}(\lceil \alpha m \rceil, m+1-\lceil \alpha m \rceil)}.
\end{aligned}$$

The last line obeys

$$\begin{aligned}
\log\left[\frac{(\alpha/3)^{\lceil \alpha m \rceil}}{\mathrm{B}(\lceil \alpha m \rceil, m+1-\lceil \alpha m \rceil)} \right] &\sim -\alpha m \log \frac{3}{\alpha} + \alpha m \log \frac{1}{\alpha} + (1-\alpha) m \log \frac{1}{1-\alpha} \\
&= -\alpha m \log 3 + (1-\alpha) m \log \frac{1}{1-\alpha}.
\end{aligned}$$

For small α, we get $-\alpha \log 3 + (1-\alpha)\log\frac{1}{1-\alpha} = -\alpha \log 3 + (1+o(1))(1-\alpha)\alpha = -(\log 3 - 1 + o(1))\alpha$. (Note that $\log 3 - 1 = 0.0986\ldots > 0$.) This immediately yields

$$\mathbb{E}\left(\log\left(\frac{1}{U_{(\lceil \alpha m \rceil)}} \right); U_{(\lceil \alpha m \rceil)} < \frac{\alpha}{3} \right) \sim \mathrm{e}^{-(\log 3 - 1 + o(1))\alpha m},$$

which implies (F.6) since $m\mathrm{e}^{-(\log 3 - 1 + o(1))\alpha m} \to 0$.

Next, we turn to show (F.7). Note that $\mathbb{P}\left(U_{(i)} \le \frac{q(k+i)}{m} \right)$ actually is the tail probability of the binomial distribution with m trials and success probability $\frac{q(k+i)}{m}$. Hence, by the Chernoff bound, this probability is bounded as

$$\mathbb{P}\left(U_{(i)} \le \frac{q(k+i)}{m} \right) \le \exp\left(-m \, \mathrm{KL}(i/m || q(k+i)/m) \right),$$

where $\mathrm{KL}(a||b) := a \log \frac{a}{b} + (1-a)\log\frac{1-a}{1-b}$ is the Kullback-Leibler divergence. Thanks to $i \ge \lceil \alpha m \rceil \gg k$, simple analysis reveals that

$$\mathrm{KL}(i/m||q(k+i)/m) \geq (1+o(1))i\left(\log\frac{1}{q} - 1 + q\right)/m.$$

Combining the last two displays gives

$$\mathbb{P}\left(U_{(i)} \leq \frac{q(k+i)}{m}\right) \leq \mathrm{e}^{-(1+o(1))\left(\log\frac{1}{q}-1+q\right)i}.$$

Plugging the above inequality into (F.7) yields

$$\sum_{i=\lceil\alpha m\rceil}^{m-k} \mathbb{P}\left(U_{(i)} \leq \frac{q(k+i)}{m}\right) \leq \sum_{i=\lceil\alpha m\rceil}^{m-k} \mathrm{e}^{-(1+o(1))\left(\log\frac{1}{q}-1+q\right)i} \to 0,$$

where the last step follows from $\log\frac{1}{q} - 1 + q > 0$ and $\lceil\alpha m\rceil \to \infty$. ■

G gFDR control for weakly correlated groups

G.1 The proof of Theorem 2.7

We will start with the auxiliary proposition, which let us to define the subgradient of convex function f by a weaker condition than in the original definition.

Proposition G.1. *For any open set H containing zero the subgradient of convex function f at b could be equivalently defined as a vector g satisfying $f(b+h) \geq f(b) + g^\mathsf{T} h,$ for all $h \in H$.*

Proof. Suppose that f is convex function and for some $b, g \in \mathbb{R}^p$ it occurs $f(b+h) \geq f(b) + g^\mathsf{T} h$ for $h \in H$, where H is an open set containing zero. Let $h_0 \in \mathbb{R}^p$ be arbitrary vector. Function $F: \mathbb{R} \to \mathbb{R}$, defined as $F(t) := f(b + th_0) - tg^\mathsf{T} h_0$, is convex. There exists $t_0 \in (0,1)$ such that $t_0 h_0 \in H$, what gives

$$f(b) \leq F(t_0) = F\big((1-t_0)\cdot 0 + t_0 \cdot 1\big) \leq (1-t_0)f(b) + t_0 F(1) \tag{G.1}$$

and $f(b+h_0) \geq f(b) + g^\mathsf{T} h_0$ as a result. ■

The proof of Theorem 2.7. Since objective function, f, in gSLOPE optimization problem is convex and subdifferentiable, the optimality condition is simply given by $0 \in \partial f(b)$, which (for all w_i's equal to w) simply leads to expression $X^\intercal(y - Xb) \in \partial J_\lambda\big(w[\![b]\!]_I\big)$. Let $b \in \mathbb{R}^p$ be such that $\|b_{I_1}\|_2 > \ldots > \|b_{I_s}\|_2 > 0$ and $\|b_{I_j}\|_2 = 0$ for $j > s$. Set $w = 1$ and suppose that g is any vector belonging to $\partial J_{\lambda,I}(b)$. Define the set $H := \big\{h \in \mathbb{R}^p : \ \|(b+h)_{I_1}\|_2 > \ldots > \|(b+h)_{I_s}\|_2, \ \|(b+h)_{I_s}\|_2 > \|h_{I_j}\|_2, \ j > s\big\}$. Since $g \in \partial J_{\lambda,I}(b)$, from the definition of subgradient for all $h \in H$ it holds

$$\sum_{i=1}^{s} \lambda_i \|(b+h)_{I_i}\|_2 + \sum_{i=s+1}^{m} \lambda_i\big([\![b+h]\!]_I\big)_{(i)} \geq \sum_{i=1}^{s} \lambda_i \|b_{I_i}\|_2 + \sum_{i=1}^{s} g_{I_i}^\intercal h_{I_i} + (g^c)^\intercal h^c, \quad \text{(G.2)}$$

for $g^c := (g_{I_{s+1}}^\intercal, \ldots, g_{I_m}^\intercal)^\intercal$ and $h^c := (h_{I_{s+1}}^\intercal, \ldots, h_{I_m}^\intercal)^\intercal$. Define $\widetilde{I} := \big\{\widetilde{I}_1, \ldots, \widetilde{I}_{m-s}\big\}$, with set $\widetilde{I}_i := \big\{(i-1)\cdot l + 1, \ldots, i \cdot l\big\}$. Then $[\![g^c]\!]_{\widetilde{I}} = [\![g]\!]_{I^c}$. Consider first case, when h belongs to the set $H^c := \{h \in H : \ h_{I_i} \equiv 0, \ i \leq s\}$. This yields $\sum_{i=1}^{m-s} \lambda_{s+i}\big([\![h^c]\!]_{\widetilde{I}}\big)_{(i)} \geq (g^c)^\intercal h^c$. Since $\{h^c : \ h \in H^c\}$ is open in $\mathbb{R}^{l(m-s)}$ and contains zero, from Proposition G.1 we have that $g^c \in \partial J_{\lambda^c,\widetilde{I}}(0)$ and this inequality is true for any $h^c \in \mathbb{R}^{l(m-s)}$ yielding

$$0 \geq \sup_{H^c} \Big\{(g^c)^\intercal h^c - J_{\lambda^c,\widetilde{I}}(h^c)\Big\} = J^*_{\lambda^c,\widetilde{I}}(g^c) = \begin{cases} 0, \ [\![g^c]\!]_{\widetilde{I}} \in C_{\lambda^c} \\ \infty, \text{otherwise} \end{cases}, \quad \text{(G.3)}$$

see Proposition C.1. This result immediately gives condition $[\![g^c]\!]_{\widetilde{I}} \in C_{\lambda^c}$, which is equivalent with $[\![g]\!]_{I^c} \in C_{\lambda^c}$. To find conditions for g_{I_i} with $i \leq s$, define sets $H_i := \{h \in H : \ h_{I_j} \equiv 0, \ j \neq i\}$. For $h \in H_i$, (G.2) reduces to $\lambda_i\|b_{I_i} + h_{I_i}\|_2 \geq \lambda_i\|b_{I_i}\|_2 + g_{I_i}^\intercal h_{I_i}$. Since the set $\{h_{I_i} : \ h \in H_i\}$ is open in $\mathbb{R}^l$ and contains zero, from Proposition G.1 (see below) we have $g_{I_i} \in \partial f_i(b_{I_i})$ for $f_i : \mathbb{R}^l \longrightarrow \mathbb{R}$, $f_i(x) := \lambda_i\|x\|_2$. Since f_i is convex and differentiable in b_{I_i}, it holds $g_{I_i} = \lambda_i \frac{b_{I_i}}{\|b_{I_i}\|_2}$. Summarizing, for any $w > 0$ we get

$$\begin{cases} g_{I_i} = w\lambda_i \frac{b_{I_i}}{\|b_{I_i}\|_2}, \ i = 1, \ldots, s \\ [\![g]\!]_{I^c} \in C_{w\lambda^c} \end{cases}, \quad \text{(G.4)}$$

which proves the left-hand side formulation in (2.22).

Observe now that, since $X_{I_i}^\intercal X_{I_i} = I_l$ for $i \leq s$, we get $v_{I_i} = X_{I_i}^\intercal\big(y - X_{\setminus I_i}\hat{\beta}_{\setminus I_i}\big) =$

$\hat{\beta}_{I_i}\Big(1+\frac{w\lambda_i}{\|\hat{\beta}_{I_i}\|_2}\Big)$. This means that, for $i \leq s$, vectors v_{I_i}'s are collinear with vectors $\hat{\beta}_{I_i}$'s. Since $1+\frac{w\lambda_i}{\|\hat{\beta}_{I_i}\|_2} > 0$, we have $\frac{v_{I_i}}{\|v_{I_i}\|_2} = \frac{\hat{\beta}_{I_i}}{\|\hat{\beta}_{I_i}\|_2}$. This yields $\hat{\beta}_{I_i} = \big(1-\frac{w\lambda_i}{\|v_{I_i}\|_2}\big)v_{I_i}$ and $\|\hat{\beta}_{I_i}\|_2 = \big|\|v_{I_i}\|_2 - w\lambda_i\big|$, what justifies the right-hand side expression in (2.22). ■

G.2 The proof of Theorem 2.8

The proof of Theorem 2.8 follows from the auxiliary Lemma, which we present below.

Lemma G.2. *Suppose that $X \in M(n,r)$, with $r+1 < n$, and entries of X are independent and identically distributed, $X_{ij} \sim \mathcal{N}(0,1/n)$ for all i and j. Let A_X and $M_{X,\lambda}$ be matrices defined as $A_X := X(X^\intercal X)^{-1}X^\intercal$ and $M_{X,\lambda} := B_X H_{\lambda,\beta} H_{\lambda,\beta}^\intercal B_X^\intercal$, for $B_X := X(X^\intercal X)^{-1}$ and $H_{\lambda,\beta}$ defined in (2.24). Then*

- ***I.1*** *there exists the expected value of A_X and $\mathbb{E}_A := \mathbb{E}\left(A_X\right) = \frac{r}{n}\mathbf{I}_n$;*
- ***I.2*** *there exists the expected value of $M_{X,\lambda}$ and $\mathbb{E}_M := \mathbb{E}\left(M_{X,\lambda}\right) = \frac{\|\lambda_S\|_2^2}{n-r-1}\mathbf{I}_n$.*

Proof. The claim ***I.1*** is obvious for $n=1$ and we will assume that $n>1$. It can be easily noticed that A_X is a symmetric, idempotent (meaning that $A_X A_X = A_X$) matrix and that $\text{trace}(A_X) = \text{trace}\left(X^\intercal X(X^\intercal X)^{-1}\right) = r$. We will now show that for each $i \in \{1,\ldots,n\}$, $j \in \{1,\ldots,r\}$ the support of a $A_X(i,j)$ distribution is bounded, which will give us the existence of the expected value. Let $\|A\|_F$ be the Frobenius norm. Then

$$\left|(A_X)_{i,j}\right| \leq \|A\|_F = \sqrt{\text{trace}(A_X^\intercal A_X)} = \sqrt{\text{trace}(A_X)} = \sqrt{r}. \tag{G.5}$$

Since entries of matrix X are randomized independently with the same distribution, $\mathbb{E}_A$ is invariant with respect to the permutation applied to rows of X, i.e. $\mathbb{E}\left(A_X\right) = \mathbb{E}\left(A_{PX}\right)$ for any permutation matrix P. This gives $\mathbb{E}_A = P\mathbb{E}_A P^\intercal$, which means that applying the same permutation to rows and columns has no impact on expected value. We will show that

$$(\mathbb{E}_A)_{i,j} = (\mathbb{E}_A)_{1,n}, \text{ for } i<j. \tag{G.6}$$

Consider first the case when $i = 1$ and $1 < j < n$. Denoting by $P_{j\leftrightarrow n}$ matrix corresponding to transposition which replaces elements j and n, we have $(\mathbb{E}_A)_{1,j} = \left(P_{j\leftrightarrow n}\mathbb{E}_A P^{\mathsf{T}}_{j\leftrightarrow n}\right)_{1,j} = (\mathbb{E}_A)_{1,n}$. When $j = n$ and $1 < i < n$, the same reasoning works with $P_{1\leftrightarrow i}$. Suppose now, that $1 < i < n$ and $1 < j < n$. We get $(\mathbb{E}_A)_{i,j} = (\mathbb{E}_A)_{1,n}$ analogously by using arbitrary permutation matrix P which replaces element j with n and element i with 1. Since $\mathbb{E}_A$ is symmetric, (G.6) is true also for $i > j$. On the other hand, for all $i, j \in \{1, \dots, n\}$, we have $(\mathbb{E}_A)_{i,i} = \left(P_{j\leftrightarrow i}\mathbb{E}_A P^{\mathsf{T}}_{j\leftrightarrow i}\right)_{i,i} = (\mathbb{E}_A)_{j,j}$. Consequently, all off-diagonal entries of $\mathbb{E}_A$ are equal to some t and all diagonal entries have the same value d. Since $nd = \operatorname{trace}(\mathbb{E}_A) = \sum_{i=1}^{n} \mathbb{E}\big(A_X(i,i)\big) = \mathbb{E}\Big(\sum_{i=1}^{n}(A_X)_{i,i}\Big) = r$, we have $d = \frac{r}{n}$ and it remains to show that $t = 0$. Define $S := \left[\begin{array}{c|c} -1 & \mathbf{0}^{\mathsf{T}} \\ \hline \mathbf{0} & \mathbf{I}_{n-1}\end{array}\right]$. Then SX differs from X only by signs of the first row. Since entries of matrix X have zero-symmetric distribution, we have $\mathbb{E}\left(A_X\right) = \mathbb{E}\left(A_{SX}\right)$. Now

$$\left[\begin{array}{c|c} d & \mathbf{1}^{\mathsf{T}}_{n-1}t \\ \hline \mathbf{1}_{n-1}t & \ddots \end{array}\right] = \mathbb{E}_A = \mathbb{E}\left(A_{SX}\right) = S\mathbb{E}_A S = \left[\begin{array}{c|c} d & -\mathbf{1}^{\mathsf{T}}_{n-1}t \\ \hline -\mathbf{1}_{n-1}t & \ddots \end{array}\right], \tag{G.7}$$

which implies $t = 0$ and proves ***I.1***.

To prove ***I.2*** observe that $M_{X,\lambda}$ is symmetric, positive semi-definite matrix. Denote by $\|M_{X,\lambda}\|_*$ the nuclear (trace) norm of matrix $M_{X,\lambda}$. We have

$$\begin{aligned}&\mathbb{E}\big|(M_{X,\lambda})_{i,j}\big| \le \mathbb{E}\big(\|M_{X,\lambda}\|_*\big) = \mathbb{E}\big(\operatorname{trace}(M_{X,\lambda})\big) = \mathbb{E}\big(\operatorname{trace}(H^{\mathsf{T}}_{\lambda,\beta}B^{\mathsf{T}}_X B_X H_{\lambda,\beta})\big) = \\ &\mathbb{E}\big(H^{\mathsf{T}}_{\lambda,\beta}(X^{\mathsf{T}}X)^{-1}H_{\lambda,\beta}\big) = \frac{n}{(n-r-1)}\,H^{\mathsf{T}}_{\lambda,\beta}H_{\lambda,\beta} = \frac{n\,\|\lambda^S\|_2^2}{n-r-1},\end{aligned} \tag{G.8}$$

since $X^{\mathsf{T}}X$ follows an inverse Wishart distribution. This gives the existence of $\mathbb{E}_M$. Analogously to situation from the proof of ***I.1***, $\mathbb{E}_M$ is invariant with respect to permutation or signs changes applied to rows of X. Since $\mathbb{E}\left(M_{PX,\lambda}\right) = P\mathbb{E}_M P^{\mathsf{T}}$ and $\mathbb{E}\left(M_{SX,\lambda}\right) = S\mathbb{E}_M S$, as before we have that $\mathbb{E}_M$ is a diagonal matrix with all diagonal entries having the same value d. The value d can be easy found by using (G.8), since we have $nd = \operatorname{trace}\big(\mathbb{E}_M\big) = \frac{n\,\|\lambda^S\|_2^2}{n-r-1}$. ■

Proof of Theorem 2.8. The claim that $\hat{v}_{I_i}$ has expected value equal to zero for each $i > s$ follows simply from the fact that $I_i \cap I_S = \emptyset$ for such i and all entries of X are assumed to be randomized independently from the normal distribution with mean zero. To find the covariance matrix we will express vector $\hat{v}_{I_i}$ in the form

$$\hat{v}_{I_i} = \overbrace{X_{I_i}^{\mathsf{T}}\Big[\mathbf{I}_n - \underbrace{X_{I_S}(X_{I_S}^{\mathsf{T}}X_{I_S})^{-1}X_{I_S}^{\mathsf{T}}}_{A_X}\Big]z}^{\xi_{X,z}} + \overbrace{wX_{I_i}^{\mathsf{T}}\underbrace{X_{I_S}(X_{I_S}^{\mathsf{T}}X_{I_S})^{-1}}_{B_X}H_{\lambda,\beta}}^{\zeta_X} \tag{G.9}$$

Since $\mathbb{E}(\xi_{X,z}\zeta_X^{\mathsf{T}}) = 0$, it holds $Cov(\hat{v}_{I_i}) = Cov(\xi_{X,z}) + Cov(\zeta_X)$. Now thanks to Lemma G.2 we have

$$\begin{aligned} Cov(\xi_{X,z}) = \mathbb{E}\left[X_{I_i}^{\mathsf{T}}\big(\mathbf{I}_n - A_X\big)zz^{\mathsf{T}}\big(\mathbf{I}_n - A_X\big)^{\mathsf{T}}X_{I_i}\right] = \\ \mathbb{E}\left[X_{I_i}^{\mathsf{T}}\big(\mathbf{I}_n - A_X\big)\big(\mathbf{I}_n - A_X\big)^{\mathsf{T}}X_{I_i}\right] = \mathbb{E}\left[X_{I_i}^{\mathsf{T}}\big(\mathbf{I}_n - A_X\big)X_{I_i}\right] = \\ \frac{1}{n}\big(n - ls\big)\cdot\mathbb{E}\left[X_{I_i}^{\mathsf{T}}X_{I_i}\right] = \frac{1}{n}\big(n - ls\big)\cdot\mathbf{I}_l, \end{aligned} \tag{G.10}$$

$$\begin{aligned} Cov(\zeta_X) = w^2\mathbb{E}\left[X_{I_i}^{\mathsf{T}}B_XH_{\lambda,\beta}H_{\lambda,\beta}^{\mathsf{T}}B_X^{\mathsf{T}}X_{I_i}\right] = w^2\frac{\|\lambda^S\|_2^2}{n - sl - 1}\mathbb{E}\left[X_{I_i}^{\mathsf{T}}X_{I_i}\right] = \\ w^2\frac{\|\lambda^S\|_2^2}{n - sl - 1}\mathbf{I}_l. \end{aligned} \tag{G.11}$$

■

H Expected maximal group effect under the total null hypothesis

Consider the case when all submatrices X_{I_i} have the same rank, $l > 0$, $w > 0$ is used as the universal weight and X is orthogonal at groups level. From the interpretation of gSLOPE estimate coming from (2.12), we see that the identification of the relevant groups could be summarized as follows: λ decides on the number, R, of groups labeled as relevant, which correspond to indices of the R largest values among $w^{-1}\|\widetilde{y}_{\mathbb{I}_1}\|_2, \ldots, w^{-1}\|\widetilde{y}_{\mathbb{I}_m}\|_2$. The random variables $w^{-1}\|\widetilde{y}_{\mathbb{I}_i}\|_2$ have a (possibly) noncentral χ distributions with l degrees of freedom and noncentrality parameters given by the entries of $[\![\widetilde{\beta}]\!]_{\mathbb{I}}$. Now, the nonzero $\|\widetilde{\beta}_{\mathbb{I}_i}\|_2$ could be perceived as a strong signal, if with high probability the random variable having the noncentral χ distribution with the noncentrality parameter $\|\widetilde{\beta}_{\mathbb{I}_i}\|_2$ is large compared to the background composed of the independent random variables with the χ_l distributions (then signal is likely to be identified by gSLOPE; otherwise, the signal could be easily covered by random disturbances and its identification has more in common with good luck than with the usage of particular method). The important quantity, which could be treated as a breaking point, is the expected value of the maximum of the background noise. Group effects being close to this value, could be perceived as medium under the orthogonal case and weak under the occurrence of correlations between groups. The above reasoning applied to the considered case, yields the issue of approximation of the expected value of the maximum of m independent χ_l-distributed variables. Suppose that $\Psi_i \sim \chi_l$ for $i = \{1, \ldots, m\}$. From Jensen's inequality we have

$$\mathbb{E}\left(\max_{i=1,\ldots,m}\{\Psi_i\}\right) = \mathbb{E}\left(\sqrt{\max_{i=1,\ldots,m}\{\Psi_i^2\}}\right) \leq \sqrt{\mathbb{E}\left(\max_{i=1,\ldots,m}\{\Psi_i^2\}\right)},$$

hence we will replace the last problem by the problem of finding a reasonable upper bound on the expected value of the maximum of m independent, χ_l^2-distributed variables.

Theorem H.1. *Let $\Psi_1, \ldots, \Psi_m$ be independent variables, $\Psi_i \sim \chi_l^2$ for all i. Then*

$$\mathbb{E}\left(\max_{i=1,\ldots,m}\{\Psi_i\}\right) \leq \frac{4\ln(m)}{1 - m^{-\frac{2}{l}}}. \tag{H.1}$$

Proof. Denote $M_m := \max_{i=1,\ldots,m}\{\Psi_i\}$. From the Jensen's inequality applied to e^{tM_m} we have

$$e^{t\mathbb{E}[M_m]} \leq \mathbb{E}\left[e^{tM_m}\right] = \mathbb{E}\left[\max_{i=1,\ldots,m} e^{t\Psi_i}\right] \leq \sum_{i=1}^{m}\mathbb{E}\left[e^{t\Psi_i}\right]. \tag{H.2}$$

We will consider only $t \in [0, \frac{1}{2})$. Since the moment generating function for χ_l^2 distribution is given by $MGF := (1-2t)^{-\frac{l}{2}}$, for each i it holds $\mathbb{E}\left[e^{t\Psi_i}\right] = (1-2t)^{-\frac{l}{2}}$ and we get $e^{t\mathbb{E}[M_m]} \leq m(1-2t)^{-\frac{l}{2}}$. Applying the natural logarithm to both sides yields

$$\mathbb{E}[M_m] \leq \frac{\ln(m) + \ln\left((1-2t)^{-\frac{l}{2}}\right)}{t}, \quad t \in [0, 1/2)\,. \tag{H.3}$$

Define $t_{m,l} := \frac{1-m^{-\frac{2}{l}}}{2}$. Then for all positive, natural numbers l and m we have $t_{m,l} \in [0, \frac{1}{2})$. Plugging $t_{m,l}$ to the right side of (H.3) gives inequality (H.1) and finishes the proof. ∎

The above theorem gives us the motivation to use the quantity $\sqrt{4\ln(m)/(1-m^{-2/l})}$ as the upper bound on the expected value of maximum over m independent χ_l-distributed variables. In all simulations, which we have performed to investigate the performance of gSLOPE, we have generated the effects for truly relevant groups basing on these upper bounds. In particular, in experiments where l_i's as well as weights were identical, we aimed at $\mathbb{E}\left(\|\widetilde{y}_{\mathbb{I}_i}\|_2\right) = \sqrt{4\ln(m)/(1-m^{-2/l})}$, for the truly relevant group i. Since $\mathbb{E}\left(\|\widetilde{y}_{\mathbb{I}_i}\|_2\right) \approx \sqrt{\|\widetilde{\beta}_{\mathbb{I}_i}\|_2^2 + l}$, this yields the setting

$$\|\widetilde{\beta}_{\mathbb{I}_i}\|_2 = B(m,l), \qquad \text{for} \qquad B(m,l) := \sqrt{4\ln(m)/(1-m^{-2/l}) - l} \tag{H.4}$$

for groups chosen to be truly relevant.

I Screening procedure for SNPs

We used the genotype and the phenotype data from the North Finland Birth Cohort (NFBC1966) dataset, described in detail in (Sabatti et al., 2009) and available in dbGaP with accession number phs000276.v2.p1 (`http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000276.v2.p1`). The raw data contains 364,590 markers for 5,402 subjects. To obtain a set of suitable weakly correlated SNPs for use in the simulation study, we screened the data as follows.

a) The genotype data was filtered in PLINK using the criteria that each SNP should be in Hardy-Weinberg equilibrium (HWE) with p-value at least 0.0001, have minor allele frequency (MAF) at least 0.01, and have a call rate of at least 95%. Also, copy number variants (CNVs) were excluded. This resulted in a screened set of 334,103 SNPs.

b) We applied the PLINK *clump* command. This command requires as input p-values for association to a phenotype. To obtain these, we performed association analysis in EMMAX using a Balding-Nichols marker-based kinship matrix to adjust for population structure (see Kang et al., 2010). As the response variable, we used the residuals for high-density lipoproteins (HDL) after regressing out the effects of sex, pregnancy or oral contraceptive use, following the analysis in (Sabatti et al., 2009).

c) PLINK *clump* was applied using a significance threshold of 0.2 for index SNPs, a physical distance window of 1 kb, and a linkage disequilibrium (LD) threshold for clumping of $r^2 = 0.1$. This corresponds to a maximum absolute correlation for nearby SNPs of $\sqrt{0.1} = 0.316$.

d) Since PLINK uses a distance-based cut-off on clumps, the resulting set of SNPs was then re-screened in *R* to ensure that the maximum r^2 between any two SNPs was 0.1. The final set includes 26,315 SNPs.

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. National Bureau of Standards, Applied Mathematics Series, vol. 55.

Bogdan, M., van den Berg, E., Sabatti, C., Su, W., and Candès, E. J. (2015). Slope – adaptive variable selection via convex optimization. *Annals of Applied Statistics*, 9(3):1103–1140.

Bogdan, M., van den Berg, E., Su, W., and Candès, E. J. (2013). Statistical estimation and testing via the ordered ℓ_1 norm. *arXiv:1310.1969*.

Donoho, D. L. and Johnstone, I. M. (1994). Minimax risk over ℓ_p-balls for ℓ_q-error. *Probability Theory and Related Fields*, 99(2):277–303.

Hardy, G., Littlewood, J. E., and Pólya, G. (1952). *Inequalities*. Cambridge University Press.

Inglot, T. (2010). Inequalities for quantiles of the chi-square distribution. *Probability and Mathematical Statistics*, 30(2):339–351.

Kang, H., Sul, J., Service, S., Zaitlen, N., Kong, S., Freimer, N., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42:348–355.

Sabatti, C., Service, S. K., Hartikainen, A., Pouta, A., Ripatti, S., Brodsky, J., Jones, C. G., Zaitlen, N. A., Varilo, T., Kaakinen, M., Sovio, U., Ruokonen, A., Laitinen, J., Jakkula, E., Coin, L., Hoggart, C., Collins, A., Turunen, H., Gabriel, S., Elliot, P., McCarthy, M. I., Daly, M. J., Jrvelin, M., Freimer, N. B., and Peltonen, L. (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature Genetics*, 41(1):35–46.

Su, W. and Candès, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Annals of Statistics*, 40:1038–1068.