1. Selection

Theorem 1.1. *If* $h\sqrt{n}b_n \xrightarrow{p} \infty$ *then* $P\left\{\hat{\boldsymbol{\beta}}_{(b)}(\boldsymbol{s}) = 0\right\} \to 1$.

Proof. The proof is by contradiction.

Recall that the objective to be minimized by $\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})$ is

$$Q\left\{\hat{\boldsymbol{\beta}}(\boldsymbol{s})\right\} = (1/2)\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\hat{\boldsymbol{\beta}}(\boldsymbol{s})\right\}^T \boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\hat{\boldsymbol{\beta}}(\boldsymbol{s})\right\} + \sum_{j=1}^{p} \lambda_j \|\hat{\boldsymbol{\beta}}_p(\boldsymbol{s})\| \tag{1}$$

Let $\hat{\boldsymbol{\beta}}_p(\boldsymbol{s}) \neq 0$. Then (1) is differentiable w.r.t. $\boldsymbol{\beta}_p(\boldsymbol{s})$ and Q is maximized at

$$\begin{aligned}
0 &= \boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}_{(-p)}(\boldsymbol{s})\hat{\boldsymbol{\beta}}_{(-p)}(\boldsymbol{s}) - \boldsymbol{Z}_{(p)}(\boldsymbol{s})\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\right\} - \lambda_p \frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\|} \\
&= \boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left[\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s}) - \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{vv}(\boldsymbol{s})\right\}\right] \\
&\qquad + \boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(-p)}(\boldsymbol{s})\left[\boldsymbol{\beta}_{(-p)}(\boldsymbol{s}) + \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{(-p),uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{(-p),vv}(\boldsymbol{s})\right\} - \hat{\boldsymbol{\beta}}_{(-p)}(\boldsymbol{s})\right] \\
&\qquad + \boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(p)}(\boldsymbol{s})\left[\boldsymbol{\beta}_{(-p)}(\boldsymbol{s}) + \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{(p),uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{(p),vv}(\boldsymbol{s})\right\} - \hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\right] \\
&\qquad - \lambda_p \frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\|}
\end{aligned} \tag{2}$$

So

$$
\begin{aligned}
\lambda_p \frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\|} &= \boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left[\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s}) - \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{vv}(\boldsymbol{s})\right\}\right] \\
&+ \left\{\boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(-p)}(\boldsymbol{s})\right\}\left[\boldsymbol{\beta}_{(-p)}(\boldsymbol{s}) + \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{(-p),uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{(-p),vv}(\boldsymbol{s})\right\} - \hat{\boldsymbol{\beta}}_{(-p)}(\boldsymbol{s})\right] \\
&+ \left\{\boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(-p)}(\boldsymbol{s})\right\}\left[\boldsymbol{\beta}_{(-p)}(\boldsymbol{s}) + \frac{h^2\kappa_2}{2\kappa_0}\left\{\boldsymbol{\beta}_{(p),uu}(\boldsymbol{s}) + \boldsymbol{\beta}_{(p),vv}(\boldsymbol{s})\right\} - \hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\right] \quad (3)
\end{aligned}
$$

From Lemma 2 of **?**, $n^{-1}\left\{\boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(-p)}(\boldsymbol{s})\right\} = O_p(1)$ and $n^{-1}\left\{\boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\boldsymbol{Z}_{(p)}(\boldsymbol{s})\right\} = O_p(1)$. From Theorem 3 of **?**, we have that $hn^{1/2}\left[\hat{\boldsymbol{\beta}}_{(-p)}(\boldsymbol{s}) - \boldsymbol{\beta}_{(-p)}(\boldsymbol{s}) - \frac{h^2\kappa_2}{2\kappa_0}\left\{\beta_{(p),uu}(\boldsymbol{s}) + \beta_{(p),vv}(\boldsymbol{s})\right\}\right] = O_p(1)$ and $hn^{1/2}\left[\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}) - \boldsymbol{\beta}_{(p)}(\boldsymbol{s}) - \frac{h^2\kappa_2}{2\kappa_0}\left\{\beta_{(p),uu}(\boldsymbol{s}) + \beta_{(p),vv}(\boldsymbol{s})\right\}\right] = O_p(1)$. So the second and third terms of the sum in (2) are $O_p(1)$. We showed in the proof of **??** that $hn^{-1/2}\boldsymbol{Z}_{(p)}^T(\boldsymbol{s})\boldsymbol{W}(\boldsymbol{s})\left\{\boldsymbol{Y} - \boldsymbol{Z}(\boldsymbol{s})\boldsymbol{\beta}(\boldsymbol{s})\right\} = O_p(1)$.

Because the first three terms of the sum in 2 are $O_p(1)$, for $\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})$ to be a solution, we must have that $hn^{1/2}\lambda_p\frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\|} = O_p(1)$.

But since by assumption $\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}_i) \neq 0$, there must be some $k \in \{1, \ldots, d_p\}$ such that $|\hat{\beta}_{(p),k}(\boldsymbol{s})| = \max\{|\hat{\beta}_{(p),k'}(\boldsymbol{s})| : 1 \leqslant k' \leqslant d_p\}$. And for this k, we have that $|\hat{\beta}_{(p),k}(\boldsymbol{s})|/\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\| \geqslant 1/\sqrt{d_p} > 0$.

Now since $\sqrt{n}b_n \to \infty$, we have that $hn^{1/2}\lambda_p\frac{\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})}{\|\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s})\|}$ is unbounded and therefore dominates the $O_p(1)$ terms of the sum in (2). So for large enough n, $\hat{\boldsymbol{\beta}}_{(p)}(\boldsymbol{s}) \neq 0$ cannot maximize Q. □

Theorem 1.2. *If* $h\sqrt{n}a_n \xrightarrow{p} 0$ *then* $P\left\{\hat{\boldsymbol{\beta}}_{(a)}(\boldsymbol{s}) \neq 0\right\} \to 1$.

Proof. Again, the proof is by contradiction.

Assume that $\hat{\boldsymbol{\beta}}_{(k)} = 0$ for some $k < p_0$. For the adaptive group lasso, the covariate group k is shrunk to zero if

$$\left\| \left\{ \boldsymbol{Z}_{(k)}^T(\boldsymbol{s}) \boldsymbol{W}(\boldsymbol{s}) \boldsymbol{Z}_{(k)}^T(\boldsymbol{s}) \right\}^{-1} \boldsymbol{Z}_{(k)}^T(\boldsymbol{s}) \boldsymbol{r}_{(k)}(\boldsymbol{s}) \right\|^2 \leqslant \frac{h^2 n \lambda^2}{\|\tilde{\boldsymbol{\beta}}_{(k)}\|^2}$$

where $\boldsymbol{r}_{(k)}(\boldsymbol{s})$ is the residual after accounting for all covariate groups except group k. That is, $\boldsymbol{r}_{(k)}(\boldsymbol{s}) = \boldsymbol{Y} - \boldsymbol{Z}_{(-k)}(\boldsymbol{s}) \boldsymbol{\beta}_{(-k)}(\boldsymbol{s})$. But $\left\| \tilde{\boldsymbol{\beta}}_{(k)} \right\| > 0$ implies that $\left\| \left\{ \boldsymbol{Z}_{(k)}^T(\boldsymbol{s}) \boldsymbol{W}(\boldsymbol{s}) \boldsymbol{Z}_{(k)}^T(\boldsymbol{s}) \right\}^{-1} \boldsymbol{Z}_{(k)}^T(\boldsymbol{s}) \boldsymbol{r}_{(k)}(\boldsymbol{s}) \right\|^2 > 0$ and $\frac{h^2 n \lambda^2}{\|\tilde{\boldsymbol{\beta}}_{(k)}\|^2} \leqslant h^2 n a_n^2 \to 0$. So

$$P \left\{ \hat{\boldsymbol{\beta}}_{(k)}(\boldsymbol{s}) \neq 0 \right\} \leqslant P \left[\left\| \left\{ \boldsymbol{Z}_{(k)}^T(\boldsymbol{s}) \boldsymbol{W}(\boldsymbol{s}) \boldsymbol{Z}_{(k)}^T(\boldsymbol{s}) \right\}^{-1} \boldsymbol{Z}_{(k)}^T(\boldsymbol{s}) \boldsymbol{r}_{(k)}(\boldsymbol{s}) \right\|^2 \leqslant h^2 n a_n^2 \right] \to 0. \tag{4}$$

$\square$