Let $F=[f_1;\dots;f_n]'$ denote a n by p matrix and for a set of indices $S\subset\{1,\dots,p\}$ we define $\mathcal{P}_S=F[S](F[S]'F[S])^{-1}F[S]'$ denote the projection matrix on the columns associated with the indices in S.

Lemma 7 (Performance of the Post-Lasso)**.** *Under Conditions AS and RF, let $\widehat{T}_l$ denote the support selected by $\widehat{\beta}_l=\widehat{\beta}_{lL}$, $\widehat{T}_l\subseteq\widehat{I}_l$, $\widetilde{m}_l=|\widehat{I}_l\setminus T_l|$, and $\widehat{\beta}_{lPL}$ be the Post-Lasso estimator based on $\widehat{I}_l$, $l=1,\dots,k_e$. Then we have*

$$\max_{l\leqslant k_e}\|D_{il}-f_i'\widehat{\beta}_{lPL}\|_{2,n}\lesssim_P \max_{l\leqslant k_e}\left\{\frac{\sqrt{(s+\widetilde{m}_l)\log(pk_e)}}{\sqrt{n\ \phi_{\min}(s+\widetilde{m}_l)}}+\|(D_l-\mathcal{P}_{\widehat{I}_l}D_l)/\sqrt{n}\|_2\right\},$$

$$\max_{1\leqslant l\leqslant k_e}\|\widehat{\Upsilon}_l(\widehat{\beta}_{lPL}-\beta_{l0})\|_1\leqslant \max_{1\leqslant l\leqslant k_e}\frac{\left(\|\widehat{\Upsilon}^0_l\|_\infty+\|\widehat{\Upsilon}_l-\widehat{\Upsilon}^0_l\|_\infty\right)\sqrt{\widetilde{m}_l+s}}{\sqrt{\phi_{\min}(\widetilde{m}_l+s)}}\|f_i'(\widehat{\beta}_{lPL}-\beta_{l0})\|_{2,n}.$$

If in addition $\lambda/n\geqslant c\|S_l\|_\infty$, and $\widehat{\Upsilon}_l$ satisfies (3.2) with $u\geqslant 1\geqslant \ell>1/c$ in the first stage for Lasso for every $l=1,\dots,k_e$, then we have

$$\max_{l\leqslant k_e}\|(D_l-\mathcal{P}_{\widehat{I}_l}D_l)/\sqrt{n}\|_2\leqslant \max_{l\leqslant k_e}\left(u+\frac{1}{c}\right)\frac{\lambda\sqrt{s}}{n\kappa^l_{c_0}}+3c_s.$$

Proof of Lemma 7. We have that $D_l-F\widehat{\beta}_{lPL}=(I-\mathcal{P}_{\widehat{I}_l})D_l-\mathcal{P}_{\widehat{I}_l}v_l$ where I is the identity operator. Therefore for every $l=1,\dots,k_e$ we have

$$\begin{aligned}\|D_l-F\widehat{\beta}_{lPL}\|_2 &\leqslant \|D_l-\mathcal{P}_{\widehat{I}_l}D_l-\mathcal{P}_{\widehat{I}_l}v_l\|_2\\ &\leqslant \|(I-\mathcal{P}_{\widehat{I}_l})D_l\|_2+\|\mathcal{P}_{\widehat{I}_l}v_l\|_2.\end{aligned}\tag{D.1}$$

Note that $\|\mathcal{P}_{\widehat{I}_l}v_l\|_2\leqslant \|(F[\widehat{I}_l]/\sqrt{n})((F[\widehat{I}_l]'F[\widehat{I}_l]/n)^{-1})\|\ \|F[\widehat{I}_l]'v_l/\sqrt{n}\|_2$ where $\|M\|$ denotes the operator norm of the matrix M. Since $\|M\|=\sqrt{\max\text{ eig}\{M'M\}}$ we have

$$\begin{aligned}\|(F[\widehat{I}_l]/\sqrt{n})((F[\widehat{I}_l]'F[\widehat{I}_l]/n)^{-1})\| &= \sqrt{\max\text{ eig}\{(F[\widehat{I}_l]'F[\widehat{I}_l]/n)^{-1}(F[\widehat{I}_l]'F[\widehat{I}_l]/n)(F[\widehat{I}_l]'F[\widehat{I}_l]/n)^{-1}\}}\\ &= \sqrt{\max\text{ eig}\{(F[\widehat{I}_l]'F[\widehat{I}_l]/n)^{-1}\}}\\ &\leqslant \sqrt{1/\phi_{\min}(s+\widetilde{m}_l)},\end{aligned}$$

where $\widetilde{m}_l=|\widehat{I}_l\setminus T_l|$, so that the last term in (D.1) satisfies

$$\|\mathcal{P}_{\widehat{I}_l}v_l\|_2\leqslant\sqrt{1/\phi_{\min}(s+\widetilde{m}_l)}\|F[\widehat{I}_l]'v_l/\sqrt{n}\|_2\leqslant\sqrt{\frac{s+\widetilde{m}_l}{\phi_{\min}(s+\widetilde{m}_l)}}\|F'v_l/\sqrt{n}\|_\infty.$$

Under Condition RF, by Lemma 5 we have

$$\max_{l=1,\dots,k_e}\|F'v_l/\sqrt{n}\|_\infty\lesssim_P\sqrt{\log(pk_e)}\max_{l\leqslant k_e,j\leqslant p}\sqrt{\mathbb{E}_n[f_{ij}^2v_{il}^2]}.$$

Note that Condition RF also implies $\max_{l\leqslant k_e,j\leqslant p}\sqrt{\mathbb{E}_n[f_{ij}^2v_{il}^2]}\lesssim_P 1$ since $\max_{l\leqslant k_e,j\leqslant p}|(\mathbb{E}_n-\bar{\mathrm{E}})[f_{ij}^2v_{il}^2]|\to_P 0$ and $\max_{l\leqslant k_e,j\leqslant p}\bar{\mathrm{E}}[f_{ij}^2v_{il}^2]\leqslant\max_{l\leqslant k_e,j\leqslant p}\bar{\mathrm{E}}[f_{ij}^2\tilde{d}_{il}^2]\lesssim 1$.

These relations yield the first result.

Letting $\delta_l = \widehat{\beta}_{lPL} - \beta_{l0}$, the statement regarding the ℓ_1-norm of the theorem follows from

$$\|\widehat{\Upsilon}_l\delta_l\|_1 \leqslant \|\widehat{\Upsilon}_l\|_\infty\|\delta_l\|_1 \leqslant \|\widehat{\Upsilon}_l\|_\infty\sqrt{\|\delta_l\|_0}\|\delta_l\|_2 \leqslant \|\widehat{\Upsilon}_l\|_\infty\sqrt{\|\delta_l\|_0}\|f_i'\delta_l\|_{2,n}/\sqrt{\phi_{\min}(\|\delta_l\|_0)},$$

and noting that $\|\delta_l\|_0 \leqslant \widetilde{m}_l + s$ and $\|\widehat{\Upsilon}_l\|_\infty \leqslant \|\widehat{\Upsilon}_l^0\|_\infty + \|\widehat{\Upsilon}_l - \widehat{\Upsilon}_l^0\|_\infty$.

The last statement follows from noting that the Lasso solution provides an upper bound to the approximation of the best model based on $\widehat{I}_l$, since $\widehat{T}_l \subseteq \widehat{I}_l$, and the application of Lemma 6. $\square$