Supplementary Materials

Supplement A: Additional Numerical Results

Distributed Computing

We compared the runtime of the proposed group LASSO projection-posterior under centralized and distributed implementations across increasing sample sizes ($n = 10000, \ldots, 20000$) with the dimensionality fixed at $p = 100$ and $s_0 = 5$ active groups, each of size 5. In the distributed version, the full dataset was partitioned into 10 shards, each processed on a separate core to compute local summary statistics ($\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$, $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$, $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}$), which were subsequently aggregated. As shown in Figure 5, this parallelization yielded a modest reduction in total runtime for larger n, primarily due to the concurrent computation of local Gram matrices. However, the subsequent posterior sampling and projection steps, which dominate the overall computational load, incur similar costs under both centralized and distributed settings. Thus, while substantial speedups are not expected for the current problem size, the distributed framework offers two key advantages: firstly, it eliminates the need to access or share the complete raw data, making it naturally suited to privacy-preserving or data-federated settings; and secondly, it ensures that the computational time is never worse than that of the centralized alternative, with potential gains when the local computations can be efficiently parallelized. We observe similar trends for the group SCAD and adaptive group LASSO projection-posteriors; the plots are not reported to avoid repetition.

For each n used in our timing study, we decomposed wall-clock time into: (i) the summary stage (computing local/central $\boldsymbol{X}^\top\boldsymbol{X}$, $\boldsymbol{X}^\top\boldsymbol{Y}$, $\boldsymbol{Y}^\top\boldsymbol{Y}$ and aggregating in the distributed case), and (ii) the sampling+projection stage (conjugate posterior sampling of $\boldsymbol{\beta}$ and sparse

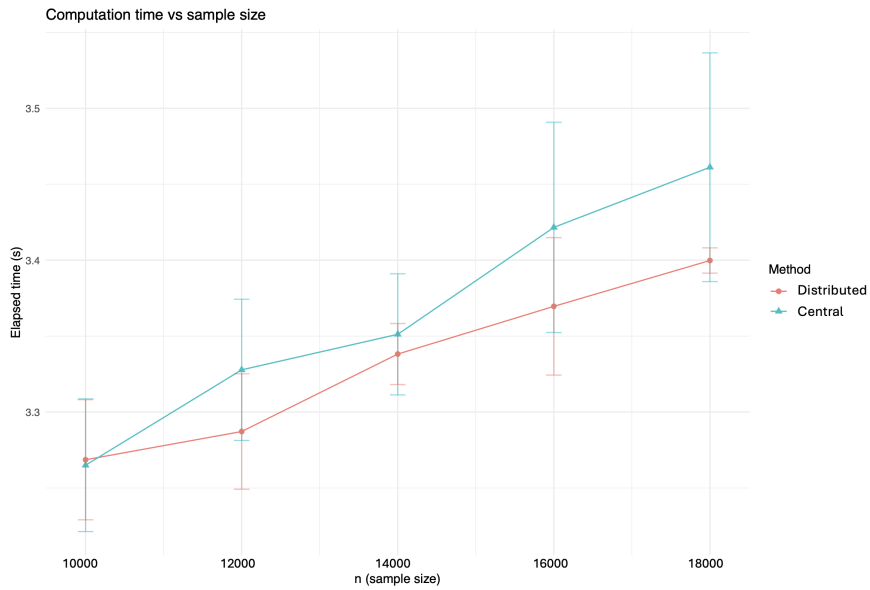

Figure 5: Computation time comparison between centralized and distributed implementations of the group LASSO projection-posterior across increasing sample sizes.

projection for $R = 1000$ draws). It should be noted that the summary step under both distributed and central computing requires an inverse of a high-dimensional matrix. In the distributed implementation with m shards, only stage (i) differs from the centralized case; stage (ii) should be similar as it mostly depends on the dimension rather than the sample size.

n	Sampling+Projection (Central)	Sampling+Projection (Dist)	Summary (Central)	Summary (Dist)	Total (Central)	Total (Dist)
10,000	2.70	2.70	0.56	0.57	3.26	3.27
14,000	2.71	2.73	0.64	0.61	3.35	3.34
18,000	2.74	2.74	0.72	0.66	3.46	3.40

Table 4: Runtime (seconds): decomposition of shared sampling + projection and summary stages under centralized vs. distributed implementations.

In our runs, the summary stage scales near-linearly with n and benefits from distributing the load, whereas the sampling+projection stage is essentially flat in n (dominated by a single

Cholesky of $\boldsymbol{X}^\top \boldsymbol{X} + a\boldsymbol{I}_p$ and R projections), hence similar under centralized and distributed setups. This explains the modest but consistent gains in Figure 5.

Let p and R be fixed. We write $T_{\text{central}}(n) = c_{\text{sum}}\, n + T_{\text{s+p}}$, $T_{\text{dist}}(n; m) = \frac{c_{\text{sum}}}{m}\, n + T_{\text{s+p}} + T_{\text{over}}$, where c_{sum} is the per-sample cost of forming summaries, $T_{\text{s+p}}$ is the (nearly n-independent) cost of sampling+projection, and T_{over} is the communication and coordination overhead cost. T_{over} is typically small and weakly dependent on n, mostly depending on the number of communication steps (one step for us) and the number of pieces to be coordinated ($m = 10$ for us). Using the empirical runtimes in Table 4, we estimate $T_{\text{s+p}} \approx 2.7$s (essentially constant across n), and $c_{\text{sum}} \approx 4 \times 10^{-5}$s per sample for the centralized case. With $m = 10$ shards and negligible T_{over}, this implies a near-linear growth in total runtime for both implementations, but with a smaller slope for the distributed version. Based on these fits, the distributed approach begins to yield visible savings beyond $n \approx 12{,}000$ (about 1-2 %), reaching roughly 8 % at $n = 50{,}000$, 14 % at $n = 100{,}000$, and 20 % at $n = 200{,}000$. Hence, while the difference is modest for the sample sizes used in our experiments, the relative gain increases linearly with n, and the distributed framework becomes substantially faster for larger datasets, while never exceeding the cost of the centralized computation.

Simulation Results for $s_0 = 20$

Below, we provide the plots for $s_0 = 20$, keeping all other parameters required for data generation fixed as in Section 5.1. We see exact similar patterns as the $s_0 = 10$ case (reported in the main paper), implying that the models are robust to varying sparsity.

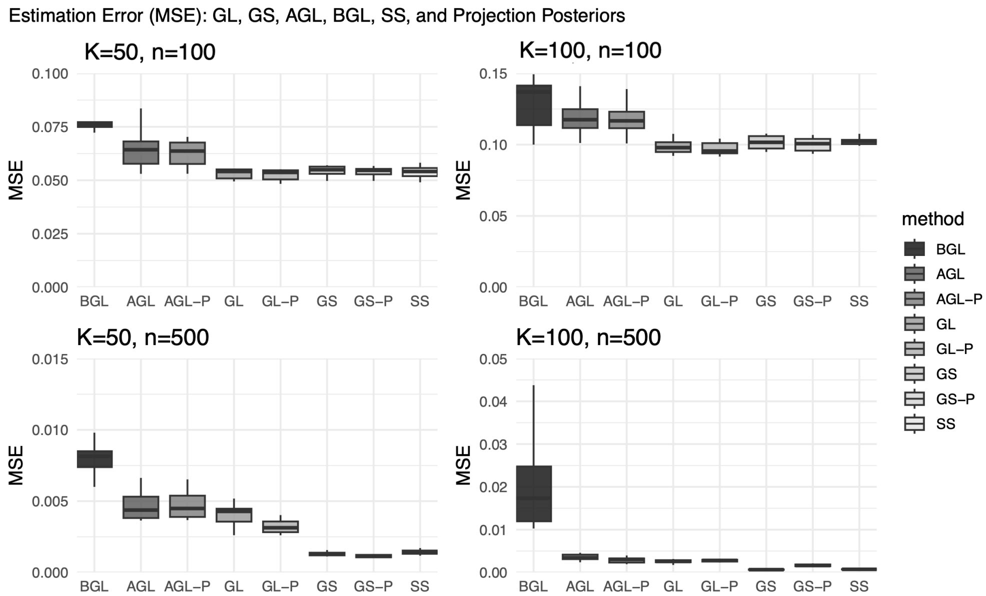

Figure 6: MSE comparisons for the four pairs of (K, n), namely (50,100),(50,500),(100,100) and (100,500) when $s_0 = 20$ groups are active, replicated 100 times

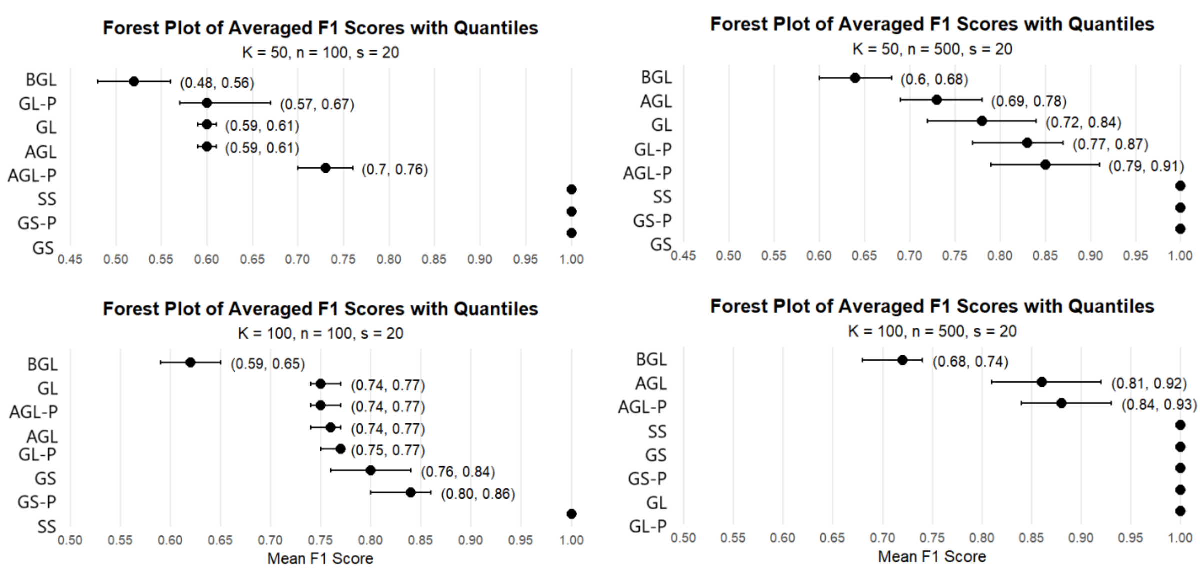

Figure 7: F1-score comparisons for the four pairs of (K, n), namely (50,100),(50,500),(100,100) and (100,500) when $s_0 = 20$ groups are active, replicated 100 times.

Setting		Debiased Group LASSO	Bayesian Group LASSO	Spike and Slab	Group ALASSO Projection	Group LASSO Projection	Group SCAD Projection	Debiased Projection
$K = 50$, $n = 100$	Signal Coverage	0.88 (0.013)	0.372 (0.059)	0.823 (0.01)	0.634 (0.027)	0.684 (0.007)	0.69 (0.005)	0.876 (0.021)
	Signal Length	1.826 (0.122)	1.9 (0.295)	2.634 (0.104)	1.03 (0.009)	0.95 (0.006)	1.47 (0.24)	1.76 (0.117)
	Noise Coverage	1	1	1	0.937 (0.016)	0.949 (0.016)	0.985 (0.013)	1
	Noise Length	1.534 (0.031)	1.84 (0.028)	3.91 (0.054)	0.522 (0.007)	0.585 (0.004)	0.704 (0.007)	1.62 (0.024)
$K = 50$, $n = 500$	Signal Coverage	0.952 (0.003)	0.717 (0.017)	0.959 (0.022)	0.962 (0.003)	0.963 (0.004)	0.959 (0.001)	0.949 (0.004)
	Signal Length	1.833 (0.009)	1.16 (0.012)	1.251 (0.012)	2.03 (0.005)	2.38 (0.007)	1.823 (0.019)	1.884 (0.027)
	Noise Coverage	1	0.969 (0.006)	1	1	1	1	1
	Noise Length	1.198 (0.03)	0.97 (0.013)	0.721 (0.006)	2.01 (0.006)	1.78 (0.007)	1.1 (0.029)	1.234 (0.015)
$K = 100$, $n = 100$	Signal Coverage	0.786 (0.034)	0.178 (0.191)	0.68 (0.059)	0.579 (0.067)	0.601 (0.054)	0.632 (0.032)	0.776 (0.03)
	Signal Length	1.207 (0.018)	0.457 (0.002)	0.81 (0.003)	0.693 (0.005)	0.782 (0.002)	0.775 (0.005)	1.218 (0.01)
	Noise Coverage	1	0.986 (0.002)	1	1	1	1	0.999 (0.0001)
	Noise Length	0.987 (0.01)	1.783 (0.012)	1.254 (0.008)	0.59 (0.011)	0.578 (0.012)	0.558 (0.013)	1.069 (0.02)
$K = 100$, $n = 500$	Signal Coverage	0.952 (0.001)	0.726 (0.103)	0.923 (0.006)	0.944 (0.003)	0.937 (0.007)	0.951 (0.002)	0.953 (0.001)
	Signal Length	1.357 (0.02)	0.659 (0.004)	0.957 (0.023)	1.297 (0.02)	1.324 (0.021)	1.109 (0.016)	1.364 (0.015)
	Noise Coverage	1	1	1	1	1	1	1
	Noise Length	0.991 (0.02)	1.645 (0.023)	0.791 (0.015)	0.61 (0.008)	0.604 (0.006)	0.598 (0.007)	0.927 (0.024)

Table 5: Average signal and noise coverage and length along with the standard error, averaged over 100 replicates when $s_0 = 20$.

Supplement B: Proof of the Theorems

First, we recall all notations introduced throughout the paper that will be used in the proofs.

Symbol	**Description**
Model, data and norms	
$\boldsymbol{\beta}, \boldsymbol{\beta}^0 \in \mathbb{R}^p$	Regression coefficients and true coefficient vector
$\\|\cdot\\|, \\|\cdot\\|_q, \\|\cdot\\|_\infty$	Euclidean/ℓ_q/sup norms
$\hat{\boldsymbol{\Sigma}} = n^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$	Gram matrix
$\boldsymbol{C}_{n(11)}$	Block of Gram on S_0
Group structure	
K	Number of groups
$G_1, \dots, G_K$	Groups (disjoint)
$p_k = \|G_k\|$	Size of the kth group
$p = \sum_{k=1}^K p_k$; $p_{\min}, p_{\max}$	Total/min/max group size
$\boldsymbol{X}_{G_k} \in \mathbb{R}^{n\times p_k}$	Sub-design for group k; $X = [X_{G_1}, \dots, X_{G_K}]$
$\boldsymbol{\beta}_k \in \mathbb{R}^{p_k}$	Group coefficient; $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\mathrm{T}}, \dots, \boldsymbol{\beta}_K^{\mathrm{T}})^{\mathrm{T}}$
$S_0 \subset \{1, \dots, K\}$	Index set of signal groups; $s_0 = \|S_0\|$
$p_0 = \sum_{k\in S_0} p_k$	Total number of active predictors
Unrestricted (conjugate) posterior	
$a_n > 0$	Prior scale; $\boldsymbol{\beta}\|\sigma \sim \mathcal{N}_p(0, \sigma^2 a_n^{-1} I_p)$
$\hat{\boldsymbol{\beta}}^{\mathrm{R}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + a_n\boldsymbol{I}_p)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$	Ridge-type posterior mean
$\boldsymbol{H}(a_n) = \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + a_n\boldsymbol{I}_p)^{-1}\boldsymbol{X}^{\mathrm{T}}$	Hat matrix; $\boldsymbol{X}\boldsymbol{\beta} \mid \sigma, \boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{R}}, \sigma^2\boldsymbol{H}(a_n))$
$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\beta}^0$	Posterior error in the mean space
	$\boldsymbol{\eta} \mid \sigma, \boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \sigma^2\boldsymbol{H}(a_n))$, $\boldsymbol{\mu} = \boldsymbol{X}(\hat{\boldsymbol{\beta}}^{\mathrm{R}} - \boldsymbol{\beta}^0)$
Projection map (sparse projection-posterior)	

(continued on next page)

Symbol	Description
$\iota : \boldsymbol{\beta} \mapsto \boldsymbol{\beta}^*$	$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{u}\in\mathbb{R}^p} \left\{ L(\boldsymbol{\beta}, \boldsymbol{u}) + \sum_{k=1}^{K} \mathcal{P}_{\lambda_n}(\\|\boldsymbol{u}_k\\|) \right\}$
$L(\boldsymbol{\beta}, \boldsymbol{u}) = n^{-1}\\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{u}\\|^2$	Loss function used in all maps
λ_n	Sparsity tuning parameter
$\tilde{\lambda}_n$	Auxiliary tuning for initial GL fit (adaptive map)
$\hat{\boldsymbol{\beta}}^{\mathrm{GL}}$	Initial group-LASSO estimator
Variance calibration (immersion posterior for σ^2)	
$\upsilon = \sigma^{-2}$	Precision; noninformative Gamma prior; map $\upsilon \mapsto$ $\upsilon^* = \kappa\,\upsilon$
$\hat{\sigma}_n^2 = n^{-1}\boldsymbol{Y}^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{H}(a_n))\boldsymbol{Y}$	Residual variance
$\kappa = \hat{\sigma}_n^2/\tilde{\sigma}^2$ for consistent $\tilde{\sigma}^2$	yields $\upsilon^*\|\boldsymbol{Y} \sim \mathrm{Gamma}\left(\frac{n}{2},\, \kappa^{-1}\frac{n\hat{\sigma}_n^2}{2}\right)$
Debiased projection and precision matrix surrogate	
$\hat{\boldsymbol{\Theta}}$	Constructed to satisfy $\hat{\boldsymbol{\Theta}}\hat{\boldsymbol{\Sigma}} \approx I_p$
$\boldsymbol{\Gamma}_g \in \mathbb{R}^{(p-p_g)\times p_g}$	Group-wise GL coefficients regressing $\boldsymbol{X}_{G_g}$ on $\boldsymbol{X}_{-g}$
$\boldsymbol{R}_g = \boldsymbol{X}_{G_g} - \boldsymbol{X}_{-g}\boldsymbol{\Gamma}_g$	Residual matrix; $\boldsymbol{T}_g^2 = n^{-1}\boldsymbol{X}_{G_g}^{\mathrm{T}}\boldsymbol{R}_g$ (block scaling)
$\boldsymbol{\Lambda}_g = \mathrm{diag}(\lambda_g^1, \ldots, \lambda_g^{p_g})$	KKT multipliers
$\boldsymbol{K}_g$	Subgradient matrix from group regressions
$\boldsymbol{\beta}^{**} = \boldsymbol{\beta}^* + \frac{1}{n}\hat{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\beta}^*)$	Debiased projection
$\boldsymbol{\Delta} = (\hat{\boldsymbol{\Theta}}\hat{\boldsymbol{\Sigma}} - \boldsymbol{I}_p)(\boldsymbol{\beta}^* - \boldsymbol{\beta}^0)$	Remainder term in debiased expansion
Rates and Constants	
r_{1n}, r_{2n}, r_{3n}	Contraction rates in $\sum_k \sqrt{p_k}\\|\beta_k^* - \beta_k^0\\|$, $\\|\beta^* - \beta^0\\|^2$, and $n^{-1}\\|X(\beta^* - \beta^0)\\|^2$
$\phi(S)$	Compatibility constant
ν	Irrepresentability margin (GL selection theory)
Additive model and splines	

(continued on next page)

Symbol	Description
f_k^0	True additive component; $Y = \sum_{k=1}^K f_k^0(X_k) + \varepsilon$
B_n	Number of B-spline basis per covariate
$\alpha > 0$	Smoothness index; typical choice $B_n \asymp n^{1/(1+2\alpha)}$
$f_k(\cdot) = \sum_{j=1}^{B_n} B_{k,j}(\cdot)\,\beta_{k,j}$	spline approximation of f_k^0; f^* induced by β^*
Credible sets and BvM	
C_j	$(1-\alpha)$ marginal credible set for β_j
$\boldsymbol{m}$, $\hat{\boldsymbol{V}}$	BvM centering and covariance for $\sqrt{n}(\beta^{**} - \beta^0)$

As the immersion posterior for σ^2 described in Section 2.2 is consistent, as in Pal and Ghosal [2024], it suffices to prove the results conditionally on given σ^*, uniformly in σ^* in a shrinking neighborhood of the true σ^0, say, $\mathcal{U}_n$, such that $\Pi(\sigma^* \in \mathcal{U}_n|\boldsymbol{Y}) \to 1$ in probability as $n \to \infty$. Since group sparsity is a special case of the general sparse high-dimensional linear regression problem, the same map used in Pal and Ghosal [2024] is reasonable.

Proof of Theorem 1(1). Since $\boldsymbol{\beta}^*$ minimizes (2.1) under the group LASSO penalty, we can write

$$\frac{1}{n}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\beta}^*\|^2 + \lambda_n \sum_{k=1}^{K} \mathcal{P}(\|\boldsymbol{\beta}_k^*\|) \leq \frac{1}{n}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\beta}^0\|^2 + \lambda_n \sum_{k=1}^{K} \mathcal{P}(\|\boldsymbol{\beta}_k^0\|). \tag{8.1}$$

Now, the left-hand side (LHS) of the above can be expressed as

$$\begin{aligned}&\frac{1}{n}\|\boldsymbol{\eta} + \boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{X}\boldsymbol{\beta}^*\|^2 + \lambda_n \sum_{k=1}^{K} \mathcal{P}(\|\boldsymbol{\beta}_k^*\|)\\ &\qquad = \frac{1}{n}\|\boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{X}\boldsymbol{\beta}^*\|^2 + \frac{1}{n}\|\boldsymbol{\eta}\|^2 - \frac{2}{n}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{\beta}^* - \boldsymbol{\beta}^0) + \lambda_n \sum_{k=1}^{K} \mathcal{P}(\|\boldsymbol{\beta}_k^*\|).\end{aligned} \tag{8.2}$$

Then, from (8.1) and (8.2), by choosing $\lambda_n \geq 2\lambda_0$, where, $\lambda_0 \asymp \sqrt{(\log p)/n}$, we can write from Lemma 6 that

$$
\begin{aligned}
&n^{-1}\|\boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{X}\boldsymbol{\beta}^*\|^2 \\
&\quad \leq \frac{2}{n}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}^*) + \lambda_n \sum_{k=1}^{K}\{\mathcal{P}(\|\boldsymbol{\beta}_k^0\|) - \mathcal{P}(\|\boldsymbol{\beta}_k^*\|)\} \qquad (8.3) \\
&\quad \leq \frac{2}{n}\max_{1\leq k\leq K}\frac{\|\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}_{G_k}\|}{\sqrt{p_k}}\sum_{k=1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k^0 - \boldsymbol{\beta}_k^*\| + \lambda_n\Big\{\sum_{k=1}^{s_0}\sqrt{p_k}(\|\boldsymbol{\beta}_k^0\| - \|\boldsymbol{\beta}_k^*\|) - \sum_{k=s_0+1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k^*\|\Big\} \\
&\quad \leq \frac{\lambda_n}{2}\Big\{\sum_{k=1}^{s_0}\sqrt{p_k}\|\boldsymbol{\beta}_k^0 - \boldsymbol{\beta}_k^*\| + \sum_{k=s_0+1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k^*\|\Big\} + \lambda_n\Big\{\sum_{k=1}^{s_0}\sqrt{p_k}\|\boldsymbol{\beta}_k^* - \boldsymbol{\beta}_k^0\| - \sum_{k=s_0+1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k^*\|\Big\} \\
&\quad = \frac{3\lambda_n}{2}\sum_{k=1}^{s_0}\sqrt{p_k}\|\boldsymbol{\beta}_k^* - \boldsymbol{\beta}_k^0\| - \frac{\lambda_n}{2}\sum_{k=s_0+1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k^*\|
\end{aligned}
$$

This then implies by Assumption 3.4,

$$
\begin{aligned}
\frac{1}{n}\|\boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{X}\boldsymbol{\beta}^*\|^2 + \frac{\lambda_n}{2}\sum_{k=1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k^* - \boldsymbol{\beta}_k^0\| &\leq 2\lambda_n\sum_{k=1}^{s_0}\sqrt{p_k}\|\boldsymbol{\beta}_k^0 - \boldsymbol{\beta}_k^*\| \\
&\leq \frac{2\lambda_n\|\boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{X}\boldsymbol{\beta}^*\|_2\sqrt{\sum_{k=1}^{s_0} p_k}}{\sqrt{n}\Phi(S_0)}. \qquad (8.4)
\end{aligned}
$$

Using $2ab \leq a^2 + b^2$ with $a = (2\sqrt{n})^{-1}\|\boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{X}\boldsymbol{\beta}^*\|_2$ and $b = (4\lambda_n/\Phi(s_0))\sqrt{\sum_{k=1}^{s_0} p_k}$ and transposing the terms leads to $\dfrac{1}{2n}\|\boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{X}\boldsymbol{\beta}^*\|^2 + \dfrac{\lambda_n}{2}\displaystyle\sum_{k=1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k^* - \boldsymbol{\beta}_k^0\| \leq \dfrac{16\lambda_n^2\sum_{k=1}^{s_0} p_k}{\Phi^2(s_0)}$.

□

Proof of Theorem 1(2). We first need to show an oracle property of the induced posterior of $\boldsymbol{\beta}^*$ obtained using the group SCAD map to prove this theorem. Consider the ideal scenario where the true model is known to us. We know exactly the set of signal groups, S_0. In such a situation, the best posterior representation would be the conjugate normal posterior but restricted to the true model S_0. If $|S_0| = s_0$, then the remaining $p - s_0$ components are set to zero, and we get the oracle projection as $\boldsymbol{\beta}_{S_0} \in \mathbb{R}^p$, such that $\boldsymbol{\beta}_{S_0}|(S = S_0, \boldsymbol{Y}, \boldsymbol{X}, \sigma^{*2}) \sim \mathcal{N}\big(\hat{\boldsymbol{\beta}}_{S_0}^{\mathrm{R}}, \sigma^{*2}\boldsymbol{H}_{S_0}(a_n)\big)$, for $\sigma^* \in \mathcal{U}_n$, where $\hat{\boldsymbol{\beta}}_{S_0}^{\mathrm{R}} \in \mathbb{R}^{s_0}$ is the ridge regression estimator with penalty a_n^{-1}, restricted to the true model and $\boldsymbol{H}_{S_0}(a_n)$ is defined in the same way as $\boldsymbol{H}(a_n)$, restricted to the truly active variables according to S_0. Let $\Pi_{S_0}(\cdot|\boldsymbol{Y})$ be used to denote the posterior distribution of the oracle projection $\boldsymbol{\beta}_{S_0}$. In the following theorem, we show that when the true value of the coefficient is larger than the specified quantity in Assumption 3.6(B), this oracle posterior itself is the minimizer of the group SCAD operation, that is, $\boldsymbol{\beta}_{S_0} = \iota(\boldsymbol{\beta}_{S_0})$, under the group SCAD penalty. This is possible because the SCAD penalty flattens out for large values of the coefficient, thereby ceasing to contribute to the minimization of the function. Thus, under a known structure, the group SCAD map returns the conjugate posterior distribution under a strong signal.

Lemma 1 (Oracle property of group SCAD projection-posterior)**.** *Under Assumptions 3.1, 3.2, 3.3 and 3.6(B), if* $\lambda_n = o\big(n^{-(1-c_3+c_4)/2}\big)$ *and* $p/(\lambda_n n^{3/2}) \to 0$, *then* $|\Pi(z|\boldsymbol{Y}) - \Pi_{S_0}(z|\boldsymbol{Y})| \to 0$ *as* $n \to \infty$.

Then, choosing $\lambda_n \geq \lambda_0$ and following the technique laid out in (8.1), (8.2) and (8.3) in the proof of Theorem 1(1), for the group SCAD penalty $n^{-1}\|\boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{X}\boldsymbol{\beta}^*\|^2 \leq 2n^{-1}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}^*) + \lambda_n \sum_{k=1}^{K}\{\mathcal{P}(\|\boldsymbol{\beta}_k^0\|) - \mathcal{P}(\|\boldsymbol{\beta}_k^*\|)\}$, which can be estimated by

$$
\begin{aligned}
&\frac{2}{n}\max_{1\leq k\leq K}\frac{\|\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}_{G_k}\|}{\sqrt{p_k}}\sum_{k=1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k^0 - \boldsymbol{\beta}_k^*\| + \lambda_n\Big\{\sum_{k=1}^{s_0}\big(\mathcal{P}(\|\boldsymbol{\beta}_k^0\|) - \mathcal{P}(\|\boldsymbol{\beta}_k^*\|)\big) - \sum_{k=s_0+1}^{K}\mathcal{P}(\|\boldsymbol{\beta}_k^*\|)\Big\}\\
\leq&\lambda_n\Big\{\sum_{k=1}^{s_0}\sqrt{p_k}\|\boldsymbol{\beta}_k^0 - \boldsymbol{\beta}_k^*\| + \sum_{k=s_0+1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k^*\|\Big\} + \frac{1}{2}(\tau+1)s_0\lambda_n^2 - \lambda_n\sum_{k=s_0+1}^{K}\mathcal{P}(\|\boldsymbol{\beta}_k^*\|)\\
=&\lambda_n\sum_{k=1}^{s_0}\sqrt{p_k}\|\boldsymbol{\beta}_k^0 - \boldsymbol{\beta}_k^*\| + \frac{1}{2}(\tau+1)s_0\lambda_n^2,
\end{aligned}
$$

by Huang and Xie [2007] and Lemma 1. Then, by Assumption 3.4, we can write

$$
\begin{aligned}
n^{-1}\|\boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{\beta}^*\|^2 + \frac{\lambda_n}{2}\sum_{k=1}^{s_0}\sqrt{p_k}\|\boldsymbol{\beta}_k^0 - \boldsymbol{\beta}_k^*\| &\leq \frac{3\lambda_n}{2}\sum_{k=1}^{s_0}\sqrt{p_k}\|\boldsymbol{\beta}_k^0 - \boldsymbol{\beta}_k^*\| + \mathcal{O}(s_0\lambda_n^2)\\
&\leq \frac{2\lambda_n\|\boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{X}\boldsymbol{\beta}^*\|\sqrt{\sum_{k=1}^{s_0}p_k}}{\sqrt{n}\Phi(S_0)} + \mathcal{O}(s_0\lambda_n^2).
\end{aligned}
$$

The rest follows as in the proof of Theorem 1(1), after (8.4). $\square$

Proof of Theorem 1(3). Let $\hat{S}$ be the set of variables selected by the group LASSO penalty. Since we weigh the adaptive LASSO map by the inverse of the group norms of $\hat{\boldsymbol{\beta}}^{\mathrm{GL}}$, the variables not included in $\hat{S}$ will receive ∞ penalty and will hence not appear in S^* too, where S^* refers to the set of variables selected by the sparse-projection (adaptive group LASSO) map. Thus, we may only restrict ourselves to $\hat{S}$ and claim that $\boldsymbol{\beta}^*_{\hat{S}}$ minimizes (2.1) under the group adaptive LASSO penalty pertaining only to the variables in $\hat{S}$. That is,

$$\frac{1}{n}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}_{\hat{S}}\boldsymbol{\beta}^*_{\hat{S}}\|^2 + \lambda_n \sum_{k\in\hat{S}} \frac{\sqrt{p_k}}{\|\hat{\boldsymbol{\beta}}_k\|}\|\boldsymbol{\beta}^*_k\| \leq \frac{1}{n}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}_{\hat{S}}\boldsymbol{\beta}^0_{\hat{S}}\|^2 + \lambda_n \sum_{k\in\hat{S}} \frac{\sqrt{p_k}}{\|\hat{\boldsymbol{\beta}}_k\|}\|\boldsymbol{\beta}^0_k\|.$$

Then, we can rewrite the above as

$$\begin{aligned}0 &\geq \frac{1}{n}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}_{\hat{S}}\boldsymbol{\beta}^0_{\hat{S}} + \boldsymbol{X}_{\hat{S}}\boldsymbol{\beta}^0_{\hat{S}} - \boldsymbol{X}_{\hat{S}}\boldsymbol{\beta}^*_{\hat{S}}\|^2 - \frac{1}{n}\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}_{\hat{S}}\boldsymbol{\beta}^0_{\hat{S}}\|^2 - \lambda_n \sum_{k\in\hat{S}} \frac{\sqrt{p_k}}{\|\hat{\boldsymbol{\beta}}_k\|}\|\boldsymbol{\beta}^*_k - \boldsymbol{\beta}^0_k\| \\ &= \frac{1}{n}\|\boldsymbol{X}_{\hat{S}}(\boldsymbol{\beta}^*_{\hat{S}} - \boldsymbol{\beta}^0_{\hat{S}})\|^2 - \frac{2}{n}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}_{\hat{S}}(\boldsymbol{\beta}^*_{\hat{S}} - \boldsymbol{\beta}^0_{\hat{S}}) - \lambda_n \sum_{k\in\hat{S}} \frac{\sqrt{p_k}}{\|\hat{\boldsymbol{\beta}}_k\|}\|\boldsymbol{\beta}^*_{\hat{S}} - \boldsymbol{\beta}^0_{\hat{S}}\|\end{aligned}$$

Now, it has been shown in Wei and Huang [2010] that there exists a positive constant ψ, such that $P(\min_{k\in S}\|\hat{\boldsymbol{\beta}}_k\| \geq \psi \min_{k\in S}\|\boldsymbol{\beta}^0_k\|) \to 1$ as $n \to \infty$. Consequently, choosing $\lambda_n \geq \lambda_0$, by Lemma 6, $n^{-1}\|\boldsymbol{X}(\boldsymbol{\beta}^* - \boldsymbol{\beta}^0)\|^2$ can be written as

$$\begin{aligned}n^{-1}\|\boldsymbol{X}_{\hat{S}}(\boldsymbol{\beta}^*_{\hat{S}} - \boldsymbol{\beta}^0_{\hat{S}})\|^2 &\leq \frac{2}{n}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}_{\hat{S}}(\boldsymbol{\beta}^*_{\hat{S}} - \boldsymbol{\beta}^0_{\hat{S}}) + \lambda_n \sum_{k\in\hat{S}} \frac{\sqrt{p_k}}{\|\hat{\boldsymbol{\beta}}_k\|}\|\boldsymbol{\beta}^*_{\hat{S}} - \boldsymbol{\beta}^0_{\hat{S}}\| \\ &\leq \frac{2}{n}\max_{1\leq k\in\hat{S}} \frac{\|\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}_{G_k}\|}{\sqrt{p_k}} \sum_{k\in\hat{S}} \sqrt{p_k}\|\boldsymbol{\beta}^0_k - \boldsymbol{\beta}^*_k\| + \frac{\lambda_n}{\psi \min_{k\in S_0}\|\boldsymbol{\beta}^0_k\|}\sum_{k\in\hat{S}} \sqrt{p_k}\|\boldsymbol{\beta}^0_k - \boldsymbol{\beta}^*_k\| \\ &\leq \lambda_n\Big(1 + \frac{1}{\psi \min_{k\in S_0}\|\boldsymbol{\beta}^0_k\|}\Big)\sum_{k\in\hat{S}} \sqrt{p_k}\|\boldsymbol{\beta}^0_k - \boldsymbol{\beta}^*_k\|\end{aligned}$$

Then, using the Compatibility condition in Assumption 3.4, we have,

$$n^{-1}\|\boldsymbol{X}(\boldsymbol{\beta}^* - \boldsymbol{\beta}^0)\|^2 \leq \lambda_n\Big(1 + \frac{1}{\psi \min_{k\in S_0}\|\boldsymbol{\beta}^0_k\|}\Big)\frac{\|\boldsymbol{X}\beta^* - \boldsymbol{X}\beta^0\|\sqrt{\sum_{k=1}^{s_0} p_k}}{\sqrt{n}\Phi(s_0)}$$

Assuming that $\min\{\|\boldsymbol{\beta}_k^0\| : k \in \hat{S}\}$ is larger than some positive constant and that the group sizes do not vary with n, we can show the following bounds holds using the inequality $2ab \leq a^2 + b^2$ and tracing the same line of argument undertaken in the proof of Theorem 1(1) succeeding (8.4), we have $n^{-1}\|\boldsymbol{X}(\boldsymbol{\beta}^* - \boldsymbol{\beta}^0)\|^2 = \mathcal{O}_P(s_0\lambda_n^2)$ and $\sum_{k=1}^K \sqrt{p_k}\|\boldsymbol{\beta}_k^* - \boldsymbol{\beta}_k^0\| = \mathcal{O}_P(s_0\lambda_n)$. □

Proof of Theorem 2(1). Defining $\boldsymbol{\xi}_S = \mathrm{vec}(\sqrt{p_k}\boldsymbol{z}_k, k \in S) \in \mathbb{R}^{\sum_{k\in S} p_k}$, we can see that the theorem will be proved if we can show that the following two hold,

$$\sup_{\sigma^* \in \mathcal{U}_n} \Pi\big(\|\boldsymbol{\beta}_{S_0}^* - \boldsymbol{\beta}_{S_0}^0\|_\infty < \min_{k \in S_0} \|\boldsymbol{\beta}_k^0\|_\infty \big| \boldsymbol{Y}, \sigma^*\big) \to 1 \text{ in probability as } n \to \infty, \tag{8.5}$$

$$\sup_{\sigma^* \in \mathcal{U}_n} \Pi\big(p_k^{-1/2}\|\boldsymbol{\xi}_k\| < 1 \quad \text{for } k \in S_0^{\mathrm{c}} \big| \boldsymbol{Y}, \sigma^*\big) \to 1 \text{ in probability as } n \to \infty, \tag{8.6}$$

From the first-order subgradient KKT condition of the group LASSO optimization problem

$$-\frac{2}{n}\boldsymbol{X}_{G_k}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) + \frac{2}{n}\boldsymbol{X}_{G_k}^{\mathrm{T}}\boldsymbol{\eta} = \lambda_n\sqrt{p_k}\boldsymbol{z}_k,$$

where $\boldsymbol{z}_k \in \mathbb{R}^{p_k}$ is such that $\|\boldsymbol{z}_k\| \leq 1$ when $\boldsymbol{\beta}_k^* = 0$ and $\boldsymbol{z}_k = \boldsymbol{\beta}_k^*/\|\boldsymbol{\beta}_k^*\|_2$ when $\boldsymbol{\beta}_k^* \neq 0$, and defining $\boldsymbol{C}_{n(11)} = n^{-1}\boldsymbol{X}_{S_0}^{\mathrm{T}}\boldsymbol{X}_{S_0}$ and $\boldsymbol{C}_{n(21)} = n^{-1}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}\boldsymbol{X}_{S_0}$, we can write

$$\boldsymbol{\beta}_{S_0}^* = \boldsymbol{\beta}_{S_0}^0 + \boldsymbol{C}_{n(11)}^{-1}\left(\frac{1}{n}\boldsymbol{X}_{S_0}^{\mathrm{T}}\boldsymbol{\eta} - \lambda_n\boldsymbol{\xi}_{S_0}\right), \tag{8.7}$$

$$\lambda_n\boldsymbol{\xi}_{S_0^{\mathrm{c}}} = \frac{1}{n}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}\boldsymbol{\eta} + \boldsymbol{C}_{n(21)}\boldsymbol{C}_{n(11)}^{-1}\left(\lambda_n\boldsymbol{\xi}_{S_0} - \frac{1}{n}\boldsymbol{X}_{S_0}^{\mathrm{T}}\boldsymbol{\eta}\right). \tag{8.8}$$

We will prove (8.5) using the condition in (8.7), and similarly, we will show (8.6) to be true using (8.8) in Lemma 2 stated below. □

Lemma 2. *Under Assumptions 3.1, 3.2, 3.3, 3.5, 3.6(A) and 3.7,*

$$\Pi\big(\|\boldsymbol{\beta}^*_{S_0} - \boldsymbol{\beta}^0_{S_0}\|_\infty < \min_{k\in S_0}\|\boldsymbol{\beta}^0_k\|_\infty \big| \boldsymbol{Y}, \sigma^*\big) \to 1 \textit{ and } \Pi\big(p_k^{-1/2}\|\boldsymbol{\xi}_k\| < 1 \quad \text{for } k \in S_0^{\mathrm{c}} \big| \boldsymbol{Y}, \sigma^*\big) \to 1$$

in probability uniformly in $\sigma^* \in \mathcal{U}_n$ *as* $n \to \infty$.

Proof of Theorem 2(2). The proof of this theorem follows from Lemma 1, which shows that the group SCAD map leads to the oracle projection $\boldsymbol{\beta}_{S_0}$. □

Proof of Theorem 2(3). With $\boldsymbol{u}_1 = \boldsymbol{u} - \boldsymbol{\beta}^0$, we can write

$$\begin{aligned}\boldsymbol{\beta}^* &= \arg\min_{\boldsymbol{u}} \left\{\frac{1}{n}\|\boldsymbol{X\beta} - \boldsymbol{X\beta}^0 + \boldsymbol{X\beta}^0 - \boldsymbol{Xu}\|^2 + \lambda_n \sum_{k=1}^K \frac{\sqrt{p_k}}{\|\hat{\boldsymbol{\beta}}_k\|}\|\boldsymbol{u}_k\|\right\} \\ &= \arg\min_{\boldsymbol{u}_1} \left\{\frac{1}{n}\|\boldsymbol{\eta} - \boldsymbol{Xu}_1\|^2 + \lambda_n \sum_{k=1}^K \frac{\sqrt{p_k}}{\|\hat{\boldsymbol{\beta}}_k\|}\|\boldsymbol{u}_1 + \boldsymbol{\beta}^0_k\|\right\},\end{aligned}$$

Defining $\boldsymbol{C}_n = n^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ and $\boldsymbol{Z}_n = n^{-1/2}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\eta}$, by the KKT conditions for $k \in S_0$,

$$\frac{1}{\sqrt{n}}\big(\boldsymbol{C}_{n(11)}(\sqrt{n}\boldsymbol{u}_{1(1)}) - \boldsymbol{Z}_{n(1)}\big)_k = -\lambda_n \frac{(\boldsymbol{u}_{1k} + \boldsymbol{\beta}^0_k)\sqrt{p_k}}{\|\boldsymbol{u}_{1k} + \boldsymbol{\beta}^0_k\|\|\hat{\boldsymbol{\beta}}_k\|} \text{ and } \|\boldsymbol{u}_{1k}\| \le \|\boldsymbol{\beta}^0_k\| \tag{a}$$

and for $k \in S_0^{\mathrm{c}}$,

$$-\lambda_n \frac{\sqrt{np_k}}{\|\hat{\boldsymbol{\beta}}_k\|}\mathbf{1}_{p_k} \le \big(\boldsymbol{C}_{n(21)}(\sqrt{n}\boldsymbol{u}_{1(1)}) - \boldsymbol{Z}_{n(2)}\big)_k \le \lambda_n \frac{\sqrt{np_k}}{\|\hat{\boldsymbol{\beta}}_k\|}\mathbf{1}_{p_k}. \tag{b}$$

Recall that $p_0 = \sum_{k=1}^{s_0} p_k$ is the total number of active variables in the model. Then, defining the vector $\boldsymbol{M}_{S_0} \in \mathbb{R}^{p_0}$ such that $\boldsymbol{M}_{S_0} = (\boldsymbol{M}_1, \ldots, \boldsymbol{M}_{s_0})^{\mathrm{T}}$, where

$$\boldsymbol{M}_k = -\lambda_n(\boldsymbol{u}_{1k} + \boldsymbol{\beta}^0_k)\sqrt{p_k}/(\|\boldsymbol{u}_{1k} + \boldsymbol{\beta}^0_k\|\|\hat{\boldsymbol{\beta}}_k\|).$$

Here, the event in (a) may be represented as

$$\begin{aligned}A =& \big\{\|(\boldsymbol{C}^{-1}_{n(11)}\boldsymbol{Z}_{n(1)} - \sqrt{n}\boldsymbol{C}^{-1}_{n(11)}\boldsymbol{M}_{S_0})_k\| = \sqrt{n}\|\boldsymbol{u}_{1k}\| \quad \text{for } k \in S_0\big\} \cap \big\{\|\boldsymbol{u}_{1k}\| \le \|\boldsymbol{\beta}^0_k\| \quad \text{for } k \in S_0\big\} \\ =& \big\{\|(\boldsymbol{C}^{-1}_{n(11)}\boldsymbol{Z}_{n(1)} - \sqrt{n}\boldsymbol{C}^{-1}_{n(11)}\boldsymbol{M}_{S_0})_k\| \le \sqrt{n}\|\boldsymbol{\beta}^0_k\| \quad \text{for } k \in S_0\big\}\end{aligned}$$

Similarly, defining $\boldsymbol{P}_{S_0} = \boldsymbol{X}_{S_0}(\boldsymbol{X}_{S_0}^{\mathrm{T}}\boldsymbol{X}_{S_0})^{-1}\boldsymbol{X}_{S_0}^{\mathrm{T}}$, the event in (b) can be simplified to

$$\begin{aligned}
B =& \big\{\|(\boldsymbol{C}_{n(21)}\big[\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{Z}_{n(1)} - \sqrt{n}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{M}_{S_0}\big] - \boldsymbol{Z}_{n(2)})_k\| \leq \lambda_n(\sqrt{np_k}/\|\hat{\boldsymbol{\beta}}_k\|)\|\mathbf{1}_{p_k}\|\big\} \\
=& \big\{\|n^{-1/2}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}\boldsymbol{P}_{S_0}\boldsymbol{\eta} - \sqrt{n}\boldsymbol{C}_{n(21)}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{M}_{S_0} - n^{-1/2}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}\boldsymbol{\eta}\| \leq \lambda_n(\sqrt{np_k}/\|\hat{\boldsymbol{\beta}}_k\|)\|\mathbf{1}_{p_k}\|\big\} \\
\subseteq& \big\{n^{-1/2}\|\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{P}_{S_0})\boldsymbol{\eta}\| \leq \lambda_n(\sqrt{np_k}/\|\hat{\boldsymbol{\beta}}_k\|)\|\mathbf{1}_{p_k}\| - \sqrt{n}\|\boldsymbol{C}_{n(21)}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{M}_{S_0}\|\big\}.
\end{aligned}$$

Consequently, for $\sigma^* \in \mathcal{U}_n$,

$$\Pi(\|\boldsymbol{\beta}_k^*\| \neq 0 \text{ for } k \in S_0, \|\boldsymbol{\beta}_k^*\| = 0 \text{ for } k \in S_0^{\mathrm{c}}|\boldsymbol{Y}) \geq \Pi(A \cap B|\boldsymbol{Y}) \geq 1 - \Pi(A^{\mathrm{c}}|\boldsymbol{Y}) - \Pi(B^{\mathrm{c}}|\boldsymbol{Y}).$$

The theorem then follows from Lemma 3 and Lemma 4. □

Lemma 3. *Under Assumptions 3.1, 3.2, 3.3 and 3.6(C),* $\sup_{\sigma^*\in\mathcal{U}_n} \Pi(A^{\mathrm{c}}|\boldsymbol{Y}, \sigma^*) \to 0$ *as* $n \to \infty$.

Lemma 4. *Under Assumptions 3.1, 3.2, 3.3 and 3.6(C),* $\sup_{\sigma^*\in\mathcal{U}_n} \Pi(B^{\mathrm{c}}|\boldsymbol{Y}, \sigma^*) \to 0$ *as* $n \to \infty$.

Proof of Theorem 3. To prove the asymptotic normality of the debiased LASSO projection $\boldsymbol{\beta}^{**}$ cemtered at the truth $\boldsymbol{\beta}^0$, we only need to control $\boldsymbol{\Delta}$, since we already know that

$$\begin{aligned}
&\sup_{\sigma^*\in\mathcal{U}_n} \mathbb{E}(n^{-1}\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\eta}|\boldsymbol{Y}, \sigma^*) = n^{-1}\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\mu}, \\
&\sup_{\sigma^*\in\mathcal{U}_n} \mathrm{var}(n^{-1}\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\eta}|\boldsymbol{Y}, \sigma^*) = n^{-1}\sigma^{*2}\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}(a_n)\boldsymbol{X}\hat{\boldsymbol{\Theta}}^{\mathrm{T}}.
\end{aligned}$$

We also know that the j-th sub-vector of the p-dimensional vector $\boldsymbol{\Delta}$ can be written as

$$\boldsymbol{\Delta}_j = (\boldsymbol{T}_j^{-2})^{\mathrm{T}}\boldsymbol{\Lambda}_j \sum_{k\neq j} \boldsymbol{K}_{j,k}^{\mathrm{T}}(\boldsymbol{\beta}_k^* - \boldsymbol{\beta}_k^0) \in \mathbb{R}^{p_j}.$$

Then, by the construction of $\hat{\boldsymbol{\Theta}}$, we have for $j = 1, 2, \ldots, K$,

$$\|\boldsymbol{\Delta}_j\| \leq \mathcal{O}\bigg(\sqrt{\big(\max_{1\leq k\leq K} p_k^2\big)\log n/n}\bigg)\sum_{k=1}^{K}\sqrt{p_k}\|\boldsymbol{\beta}_k^* - \boldsymbol{\beta}_k^0\| \leq C\lambda_0 s_0\sqrt{\big(\max_{1\leq k\leq K} p_k^3\big)\log n/n}.$$

in view of Lemma 10 and Theorem 1(1). □

Proof of Corollary 1. First, we show that the means of the two limiting normal distributions are the same. That is, we need to show that the mean of the random variable $\sqrt{n}(\boldsymbol{\beta}^{**} - \hat{\boldsymbol{\beta}}^{\text{DGL}}) = \sqrt{n}(\boldsymbol{m} + \boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}^{\text{DGL}})$ vanishes as n grows to ∞. From (3.7), we can write the above quantity as

$$\begin{aligned}\|\sqrt{n}(\boldsymbol{m} - \hat{\boldsymbol{\beta}}^{\text{DGL}} + \boldsymbol{\beta}^0)\| =& \sqrt{n}\|\hat{\boldsymbol{\Omega}}^{-1/2} n^{-1} \hat{\boldsymbol{\Theta}} \boldsymbol{X}^{\text{T}} (\boldsymbol{\mu} - \boldsymbol{\varepsilon}) + \boldsymbol{\Delta}^{\text{DGL}}\| \\ =& \sqrt{n}\|n^{-1} \hat{\boldsymbol{\Theta}} \boldsymbol{X}^{\text{T}} (\boldsymbol{X}\hat{\boldsymbol{\beta}}^{\text{R}} - \boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\beta}^0) + \boldsymbol{\Delta}^{\text{DGL}}\| \\ \leq& n^{-1/2}\|\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\text{R}}\| \cdot \|\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\text{T}}\|_{\text{op}} + \sqrt{n}\|\boldsymbol{\Delta}^{\text{DGL}}\|,\end{aligned}$$

which is $o_P(1)$ because the first term is a product of two quantities, $n^{-1/2}\|\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\text{R}}\|$ and $\|\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\text{T}}\|_{\text{op}}$, of which the first vanishes by Lemma 8 and the second can be shown to be bounded using the Proposition 5 in Honda [2021], which says that $n^{-1}\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\text{T}}\boldsymbol{X}\hat{\boldsymbol{\Theta}}^{\text{T}} \to \boldsymbol{\Theta}$ blockwise in probability. The second term $\sqrt{n}\|\boldsymbol{\Delta}^{\text{DGL}}\|$ has been proven to be $o(1)$ in Proposition 3 of Honda [2021]. For completeness and ease of understanding, both these propositions have been restated in Lemma 9 and Lemma 10, respectively. Finally, since

$$\left|\frac{\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\text{T}}\boldsymbol{H}(a_n)\boldsymbol{X}\hat{\boldsymbol{\Theta}}^{\text{T}}}{\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\text{T}}\boldsymbol{X}\hat{\boldsymbol{\Theta}}^{\text{T}}} - 1\right| = \left|\frac{\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\text{T}}\big(\boldsymbol{I}_p - \boldsymbol{H}(a_n)\big)\boldsymbol{X}\hat{\boldsymbol{\Theta}}^{\text{T}}}{\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\text{T}}\boldsymbol{X}\hat{\boldsymbol{\Theta}}^{\text{T}}}\right| \leq \operatorname{tr}\left(\boldsymbol{I}_p - \boldsymbol{H}(a_n)\right) = o(1),$$

the variances of the two asymptotic normal distributions match as $n \to \infty$. □

Proof of Theorem 4(1). Recalling that $f_k(X_{k,i}) = \sum_{j=1}^{B_n} \mathcal{B}_{k,j}(X_{k,i})\beta_{k,j}$ and the corresponding map-induced version $f_k^*(X_{k,i}) = \sum_{j=1}^{B_n} \mathcal{B}_{k,j}(X_{k,i})\beta_{k,j}^*$, we can write

$$\|f^* - f^0\|^2 \leq 2(n^{-1}\|\mathcal{B}\boldsymbol{\beta}^* - \mathcal{B}\boldsymbol{\beta}^0\|^2 + \|\mathcal{B}\boldsymbol{\beta}^0 - f_0\|^2) = \mathcal{O}(s_0 B_n \lambda_n^2) + \mathcal{O}(s_0 B_n^{-2\alpha}),$$

where the first bound follows from theorem Theorem 1(1) and the second bound is due to approximation by B-spline basis expansion of $f_0(\cdot)$. Then, for $B_n \asymp n^{1/(2\alpha+1)}$ and $\lambda_n \asymp \sqrt{(\log p)/n}$, we have $\|f^* - f^0\|^2 = \mathcal{O}(s_0 n^{-2\alpha/(2\alpha+1)} \log p)$. □

Proof of Theorem 4(2). We note that, although the true underlying data-generating model is a nonparametric additive model, we use a B-spline basis expansion followed by a group-sparsity operation to construct the model. Hence, we need to account not only for random error but also for the approximation error introduced by misspecifying the model as a B-spline regression. Thus, besides $\boldsymbol{\varepsilon}$, the model error now also includes $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^{\mathrm{T}}$, where $\delta_i = \sum_{k=1}^{K} \big(f_k^0(X_{k,i}) - f_k(X_{k,i})\big)$. We know that $\|\boldsymbol{\delta}\| = \mathcal{O}\big(s_0 n^{1/(4\alpha+2)}\big)$. The theorem then readily follows from Theorem 2(1), replacing $\boldsymbol{\varepsilon}$ by $\boldsymbol{\varepsilon}^{(2)} = \boldsymbol{\varepsilon} + \boldsymbol{\delta}$. □

Proofs of the Lemmas

Lemma 5. *Under Assumption 3.5, we have that*

$$F_n = \frac{(\boldsymbol{\eta} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{X}_{G_k} \boldsymbol{X}_{G_k}^{\mathrm{T}} (\boldsymbol{\eta} - \boldsymbol{\mu})}{(\boldsymbol{\eta} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{X}_{G_k} (\boldsymbol{X}_{G_k}^{\mathrm{T}} \boldsymbol{H}(a_n) \boldsymbol{X}_{G_k})^{-1} \boldsymbol{X}_{G_k}^{\mathrm{T}} (\boldsymbol{\eta} - \boldsymbol{\mu})} = \mathcal{O}(1). \tag{8.3}$$

Proof. It is easy to see that,

$$0 \le F_n \le \frac{\|(\boldsymbol{\eta} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{X}_{G_k}\|^2 \lambda_{\max}(\boldsymbol{X}_{G_k}^{\mathrm{T}} \boldsymbol{H}(a_n) \boldsymbol{X}_{G_k})}{\|(\boldsymbol{\eta} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{X}_{G_k}\|^2} \le \lambda_{\max}(\boldsymbol{H}(a_n)) \cdot \lambda_{\max}(\boldsymbol{X}_{G_k}^{\mathrm{T}} \boldsymbol{X}_{G_k})$$

is bounded. □

Lemma 6. *For all $x > 0$ and for*

$$\lambda_0^2(x) = \frac{8}{n} \left(1 + \sqrt{\frac{4x + 4\log p}{p_{\min}}} + \frac{4x + 4\log p}{p_{\min}}\right),$$

we have,

$$\sup_{\sigma^* \in \mathcal{U}_n} \Pi\Big(\Big\{ \max_{1 \le k \le K} \frac{\|\boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{X}_{G_k}\|^2}{n p_k} \le \frac{\lambda_0^2(x)}{4} \Big\} \Big| \boldsymbol{Y}, \sigma^* \Big) \ge 1 - \exp(-x).$$

Proof. From the posterior distribution of $\boldsymbol{X\beta}$, we can write

$$\boldsymbol{X}_{G_k}^{\mathrm{T}} (\boldsymbol{\eta} - \boldsymbol{\mu}) | (\boldsymbol{Y}, \sigma^*) \sim \mathcal{N}_{p_k}(\mathbf{0}_{p_k}, \sigma^{*2} \boldsymbol{X}_{G_k}^{\mathrm{T}} \boldsymbol{H}(a_n) \boldsymbol{X}_{G_k}).$$

Now consider the singular value decomposition of the design matrix given by $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}$, where the columns of $\boldsymbol{U}$ are the left singular vectors and $\boldsymbol{V}^{\mathrm{T}}$ has rows that are the right singular vectors, such that $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{I}_n$ and $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{I}_p$. Then, we have

$$\boldsymbol{I}_n - \boldsymbol{H}(a_n) = \boldsymbol{U}\left(\boldsymbol{I}_n - \boldsymbol{D}(\boldsymbol{D}\boldsymbol{D}^{\mathrm{T}} + a_n\boldsymbol{I}_n)^{-1}\boldsymbol{D}\right)\boldsymbol{U}^{\mathrm{T}},$$

and consequently $\boldsymbol{I}_n - \boldsymbol{H}(a_n)$ is non-negative definite. Now, note that,

$$\begin{aligned}
\max_{1\le k\le K} \|\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{X}_{G_k}\| &= \max_{1\le k\le K} \|\boldsymbol{X}_{G_k}{}^{\mathrm{T}}\boldsymbol{H}(a_n)\big(\boldsymbol{X}\boldsymbol{\theta}^0 + \boldsymbol{\varepsilon}\big) - \boldsymbol{X}_{G_k}{}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{\theta}^0\| \\
&= \max_k \|\boldsymbol{X}_{G_k}{}^{\mathrm{T}}\boldsymbol{H}(a_n)\boldsymbol{\varepsilon} - \boldsymbol{X}_{G_k}{}^{\mathrm{T}}\big(\boldsymbol{I}_n - \boldsymbol{H}(a_n)\big)\boldsymbol{X}\boldsymbol{\theta}^0\| \\
&\le \max_k \|\boldsymbol{X}_{G_k}{}^{\mathrm{T}}\boldsymbol{H}(a_n)\boldsymbol{\varepsilon}\| + \max_k \|\boldsymbol{X}_{G_k}{}^{\mathrm{T}}\big(\boldsymbol{I}_n - \boldsymbol{H}(a_n)\big)\boldsymbol{X}\boldsymbol{\theta}^0\| \\
&\le \max_k \big[\max_{1\le i\le n} |\varepsilon_i| \|\boldsymbol{X}_{G_k}{}^{\mathrm{T}}\boldsymbol{H}(a_n)\| + \max_{1\le i\le n} |\mathbb{E}(Y_i)| \|\boldsymbol{X}_{G_k}{}^{\mathrm{T}}\big(\boldsymbol{I}_n - \boldsymbol{H}(a_n)\big)\|\big] \\
&= \mathcal{O}\big(K\sqrt{\log n}\big).
\end{aligned}$$

Then, defining $\chi_k^2 = (\boldsymbol{\eta} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{X}_{G_k}(\boldsymbol{X}_{G_k}^{\mathrm{T}}\boldsymbol{H}(a_n)\boldsymbol{X}_{G_k})^{-1}\boldsymbol{X}_{G_k}^{\mathrm{T}}(\boldsymbol{\eta} - \boldsymbol{\mu})$, and noting that the posterior distribution of χ_k^2 given data follows a χ^2-distribution with p_k degrees of freedom, we get by Cauchy-Schwarz inequality, uniformly in $\sigma^* \in \mathcal{U}_n$,

$$\begin{aligned}
&\sup_{\sigma^*\in\mathcal{U}_n} \Pi\Big(\big\{\max_{1\le k\le K} \frac{\|\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}_{G_k}\|^2}{np_k} \le \lambda_0^2(x)/4\big\}\big|\boldsymbol{Y}, \sigma^*\Big) \\
\ge &\sup_{\sigma^*\in\mathcal{U}_n} \Pi\Big(\big\{\max_{1\le k\le K} \frac{\|(\boldsymbol{\eta} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{X}_{G_k}\|^2}{np_k} + \max_{1\le k\le K} \frac{\|\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{X}_{G_k}\|^2}{np_k} \le \frac{\lambda_0^2(x)}{8}\big\}\big|\boldsymbol{Y}, \sigma^*\Big) \\
\ge &1 - \sup_{\sigma^*\in\mathcal{U}_n} \Pi\Big(\big\{\max_{1\le k\le K} \frac{\chi_k^2}{np_k} \cdot F_n + \max_{1\le k\le K} \frac{\|\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{X}_{G_k}\|^2}{np_k} > \frac{\lambda_0^2(x)}{8}\big\}\big|\boldsymbol{Y}, \sigma^*\Big) \\
\ge &1 - \sup_{\sigma^*\in\mathcal{U}_n} \Pi\Bigg(\max_{1\le k\le K} \chi_k^2 \cdot F_n + \max_{1\le k\le K} \|\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{X}_{G_k}\|^2 > p_k\Bigg(1 + \sqrt{\frac{4x + 4\log p}{p_{\min}}} + \frac{4x + 4\log p}{p_{\min}}\Bigg)\Bigg|\boldsymbol{Y}, \sigma^*\Bigg) \\
\ge &1 - \exp(-x),
\end{aligned}$$

because F_n, introduced in (8.3), is bounded by Lemma 5, $\max_{1\le k\le K} \|\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{X}_{G_k}\|^2$ has been shown to be $\mathcal{O}\big(K^2 \log n\big)$ and $\mathbb{P}\big(\chi_k^2 \ge p_k(1 + a)\big) \le \exp\big(-p_k\big(a - \log(1 + a)\big)/2\big)$ by Wallace

[1959]. Then, taking $a = \sqrt{4x/p_k} + 4x/p_k$, and using the estimate $a - \log(1+a) \geq a^2/2(1+a) \geq 2x/p_k$, it follows as in Bühlmann and van de Geer [2011] that $\sup_{\sigma^* \in \mathcal{U}_n} \Pi\big(\chi_k^2 \geq p_k(1+a)\big|\boldsymbol{Y}, \sigma^*\big) \leq e^{-x}$. □

Lemma 7 (Lemma 1 of Pal and Ghosal [2024])**.** *Under Assumption 3.3,*

$$\max_{j=1,\ldots,n} \Big|1 - \frac{d_j^2}{d_j^2 + a_n}\Big| = \max_{j=1,\ldots,n} \Big|\frac{a_n}{d_j^2 + a_n}\Big| = o(n^{-1}),$$

where $d_1, \ldots, d_n$ are the singular values of $\boldsymbol{X}$.

Lemma 8 (Lemma 5 of Pal and Ghosal [2024])**.** *Under Assumption 3.3, $n^{-1/2}\|\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\theta}}^{\mathrm{R}}\|_1 = o_P(1)$.*

Lemma 9 (Proposition 3 of Honda [2021])**.** *Let m be a fixed positive integer, and $\{j_1, \ldots, j_m\} \subset \{1, 2, \ldots, K\}$. If Assumption 3.8 holds, then, uniformly in $\{j_1, \ldots, j_m\}$, as $n \to \infty$, we have that $\max(|\lambda_{min}(\boldsymbol{D})|, |\lambda_{max}|(\boldsymbol{D})) \to 0$ in true probability, where $\hat{\boldsymbol{\Omega}} = n^{-1}\hat{\boldsymbol{\Theta}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\hat{\boldsymbol{\Theta}}^{\mathrm{T}}$ and*

$$\boldsymbol{D} = \begin{bmatrix} \hat{\boldsymbol{\Omega}}_{j_1,j_1} & \ldots & \hat{\boldsymbol{\Omega}}_{j_1,j_m} \\ \vdots & \ddots & \vdots \\ \hat{\boldsymbol{\Omega}}_{j_m,j_1} & \ldots & \hat{\boldsymbol{\Omega}}_{j_m,j_m} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Theta}_{j_1,j_1} & \ldots & \boldsymbol{\Theta}_{j_1,j_m} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Theta}_{j_m,j_1} & \ldots & \boldsymbol{\Theta}_{j_m,j_m} \end{bmatrix}.$$

Lemma 10 (Proposition 5 of Honda [2021])**.** *Under Assumption 3.8 we have*

$$\|\boldsymbol{\Delta}_j^{DGL}\| < C \max_{1 \leq k \leq K} p_k^2 \cdot \frac{s_0 \log n}{n}$$

uniformly in $j = 1, 2, \ldots, K$, with probability tending to 1 for sufficiently large C.

Proof of Lemma 1. We need to show that $\boldsymbol{\beta}_{S_0}$ is the minimizer of (2.1) under the group SCAD penalty. Writing $Q = n^{-1}\|\boldsymbol{X}\boldsymbol{\beta}_{S_0} - \boldsymbol{X}\boldsymbol{u}\|^2 + \sum_{k=1}^K \mathcal{P}_{\lambda_n}(\|\boldsymbol{u}_k\|)$, for the k-th group,

$$\frac{\partial Q}{\partial \boldsymbol{u}_k} = -\frac{1}{n}\boldsymbol{X}_{G_k}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{\beta}_{S_0} - \boldsymbol{X}\boldsymbol{u}) + \frac{P'_{\lambda_n}(\|\boldsymbol{u}_k\|)\boldsymbol{u}_k}{\|\boldsymbol{u}_k\|},$$

where $P'_{\lambda_n}(t) = \lambda_n\big(\mathbb{1}(t < \lambda_n) + \lambda_n^{-1}(\tau-1)^{-1}(\tau\lambda_n - t)_+\mathbb{1}(t > \lambda_n)\big)$. Now, following Guo et al. [2015], for the oracle posterior $\boldsymbol{\beta}_{S_0}$ to be the minimizer of Q, the following second-order sufficiency conditions should be satisfied with high posterior probability:

1. $V_k(\boldsymbol{\beta}_{S_0}) = 0, \|\boldsymbol{\beta}_{S_0,k}\| > \tau\lambda_n$ for $k \leq s_0$,

2. $\|V_k(\boldsymbol{\beta}_{S_0})\| < \lambda_n, \|\boldsymbol{\beta}_{S_0,k}\| < \lambda_n$ for $k > s_0$,

assuming, for the convenience of notation, that the first s_0 groups are the active groups. The first condition holds if $\Pi(\|\boldsymbol{\beta}_{S_0,k}\| > \tau\lambda_n|\boldsymbol{Y}) \to 1$ as $n \to \infty$ for $k \leq s_0$. By the triangle inequality we have $\|\boldsymbol{\beta}_{S_0,k}\| \geq \|\boldsymbol{\beta}^0_k\| - \|\boldsymbol{\beta}_{S_0,k} - \boldsymbol{\beta}^0_k\|$.

Then by invoking Assumption 3.6(B) and recalling that $\lambda_n = o\big(n^{-(1-c_3+c_4)/2}\big)$, we observe that for any $\epsilon > 0$, the first condition boils down to

$$\sup_{\sigma^*\in\mathcal{U}_n} \Pi\big(\cup_{k\leq s_0}\left\{\sqrt{n}\|\boldsymbol{\beta}_{S_0,k} - \boldsymbol{\beta}^0_k\| \geq \epsilon n^{c_3/2}\right\}|\boldsymbol{Y},\sigma^*\big) \leq \sum_{k=1}^{s_0} \frac{\sup_{\sigma^*\in\mathcal{U}_n}\mathbb{E}\big(\sqrt{n}\|\boldsymbol{\beta}_{S_0,k} - \boldsymbol{\beta}^0_k\|\big|\boldsymbol{Y},\sigma^*\big)}{\epsilon n^{c_3/2}},$$

and can be bounded by $\epsilon^{-1}n^{-c_3/2}\sum_{k=1}^{s_0}\left\{\mathbb{E}\big(\sqrt{n}\|\boldsymbol{\beta}_{S_0,k} - \hat{\boldsymbol{\beta}}^{\mathrm{R}}_{S_0;k}\|\big|\boldsymbol{Y}\big) + \mathbb{E}\big(\sqrt{n}\|\hat{\boldsymbol{\beta}}^{\mathrm{R}}_{S_0;k} - \boldsymbol{\beta}^0_k\|\big|\boldsymbol{Y}\big)\right\} = o(1)$.

To prove the second condition, we only need $\sup_{\sigma^*\in\mathcal{U}_n}\Pi(\max_{k>s_0}\|V_k(\boldsymbol{\beta}_{S_0,k}\| < \lambda_n\big|\boldsymbol{Y},\sigma^*) \to 1$ as $n \to \infty$, the rest is trivial for $k > s_0$. First, we denote $\boldsymbol{\beta}_{S_0,(1)}$ as the vector of the components of the oracle projection $\boldsymbol{\beta}_{S_0}$ pertaining to the s_0 active groups only. Then, $\boldsymbol{\beta}_{S_0,(1)} \in \mathbb{R}^{\sum_{k\leq s_0}p_k}$. Noting that $\boldsymbol{\beta}_{S_0,(1)}|\boldsymbol{Y},\sigma^* \sim \mathcal{N}\big((\boldsymbol{X}^{\mathrm{T}}_{(1)}\boldsymbol{X}_{(1)} + a_n\boldsymbol{I}_n)^{-1}\boldsymbol{X}^{\mathrm{T}}_{(1)}\boldsymbol{Y}, \sigma^{*2}\boldsymbol{H}_{(1)}(a_n)\big)$, where $\boldsymbol{H}_{(1)}(a_n) = \boldsymbol{X}_{(1)}(\boldsymbol{X}^{\mathrm{T}}_{(1)}\boldsymbol{X}_{(1)} + a_n\boldsymbol{I}_n)^{-1}\boldsymbol{X}^{\mathrm{T}}_{(1)}$, by Assumption 3.3, we estimate the probability by

$$
\begin{aligned}
&\sum_{k>s_0}\sup_{\sigma^*\in\mathcal{U}_n}\Pi(n^{-1/2}\|\boldsymbol{X}_{G_k}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{X}_{(1)}\boldsymbol{\beta}_{S_0,(1)})\|>\sqrt{n}\lambda_n\big|\boldsymbol{Y},\sigma^*)\\
=&(p-s_0)\mathcal{O}\bigg(\frac{1}{\sqrt{n}\lambda_n\|\boldsymbol{X}_{G_k}^{\mathrm{T}}[\boldsymbol{H}(a_n)+\boldsymbol{H}_{(1)}(a_n)]\boldsymbol{X}_{G_k}\|\sigma^2}\bigg)\\
\leq&(p-s_0)\mathcal{O}\bigg(\frac{1}{\sqrt{n}\lambda_n\|\boldsymbol{X}_{G_k}^{\mathrm{T}}\boldsymbol{X}_{G_k}\|\sigma^2}\bigg),
\end{aligned}
$$

by the non-negative definiteness of $\boldsymbol{I}_n-\boldsymbol{H}(a_n)$. We know $\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{X}_{(1)}\boldsymbol{\beta}_{S_0,(1)}\big|(\boldsymbol{Y},\sigma^*)\sim\mathcal{N}_n\big(\boldsymbol{0},[\boldsymbol{H}(a_n)+\boldsymbol{H}_{(1)}(a_n)]\sigma^{*2}\big)$. Hence, the assertion follows using the lower bound for λ_n from Lemma 7. □

Proof of Lemma 2. Using Markov inequality, we have,

$$
\begin{aligned}
&\sup_{\sigma^*\in\mathcal{U}_n}\Pi\big(\|\boldsymbol{\beta}_{S_0}^*-\boldsymbol{\beta}_{S_0}^0\|_\infty>\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|_\infty\big|\boldsymbol{Y},\sigma^*\big)\\
&\quad\leq\frac{\sup_{\sigma^*\in\mathcal{U}_n}\mathbb{E}\big(\|\boldsymbol{\beta}_{S_0}^*-\boldsymbol{\beta}_{S_0}^0\|_\infty\big|\boldsymbol{Y},\sigma^*\big)}{\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|_\infty}\\
&\quad\leq\frac{1}{\alpha}\bigg[\sup_{\sigma^*\in\mathcal{U}_n}\mathbb{E}\big(\|n^{-1}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}\boldsymbol{\eta}\|_\infty\big|\boldsymbol{Y},\sigma^*\big)+\big\|\lambda_n\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{\xi}_{S_0}\big\|_\infty\bigg]\\
&\quad=\mathrm{I}+\mathrm{II},\ \text{say}.
\end{aligned}
$$

Now,

$$
\begin{aligned}
\mathrm{I}&=\frac{1}{\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|_\infty}\sup_{\sigma^*\in\mathcal{U}_n}\mathbb{E}\big(\|n^{-1}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}(\boldsymbol{\eta}-\boldsymbol{\mu})+n^{-1}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{S_0}^{\mathrm{T}}\boldsymbol{\mu}\|_\infty\big|\boldsymbol{Y},\sigma^*\big)\\
&\leq\frac{1}{\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|_\infty}\big[\sup_{\sigma^*\in\mathcal{U}_n}\mathbb{E}\big(\|n^{-1}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}(\boldsymbol{\eta}-\boldsymbol{\mu})\|_\infty\big|\boldsymbol{Y},\sigma^*\big)+\|n^{-1}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}\boldsymbol{X}(\hat{\boldsymbol{\beta}}^{\mathrm{R}}-\boldsymbol{\beta}^0)\|_\infty\big]\\
&=\frac{1}{\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|_\infty}\big[\mathcal{O}\big(\sqrt{\log p_0}/n\big)+\frac{1}{n}\|\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}\boldsymbol{X}\big((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}+a_n\mathrm{I}_p)^{-1}\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}\beta^0+\boldsymbol{\varepsilon})-\boldsymbol{\beta}^0\big)\|_\infty\big]\\
&\leq\frac{1}{\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|_\infty}\big[\mathcal{O}\big(\sqrt{\log p_0}/n\big)+\frac{1}{n}\big(\big\|\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}\big(\boldsymbol{I}_n-\boldsymbol{H}(a_n)\big)\boldsymbol{X}\boldsymbol{\beta}^0\big\|_\infty+\|\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}\boldsymbol{H}(a_n)\boldsymbol{\varepsilon}\|_\infty\big)\big]\\
&=\frac{1}{\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|_\infty}\mathcal{O}\big(\sqrt{\log p_0}/n\big),
\end{aligned}
$$

since we have that $\|\boldsymbol{X}_{S_0^c}^{\mathrm{T}}\boldsymbol{H}(a_n)\boldsymbol{\varepsilon}\|_\infty = \max_{1\le i\le n}|\varepsilon_i|\|\boldsymbol{X}_{S_0^c}^{\mathrm{T}}\boldsymbol{H}(a_n)\|_\infty$ and $\|\boldsymbol{X}_{S_0^c}^{\mathrm{T}}\big(\boldsymbol{I}_n - \boldsymbol{H}(a_n)\big)\boldsymbol{X}\boldsymbol{\beta}^0\|_\infty = \max_{1\le i\le n}|\mathbb{E}(Y_i)|\|\boldsymbol{X}_{S_0^c}^{\mathrm{T}}\big(\boldsymbol{I}_n - \boldsymbol{H}(a_n)\|_\infty$. Moreover, following Nardi and Rinaldo [2008], we can bound II by

$$\lambda_n \frac{\sqrt{\sum_{k=1}^{s_0} p_k}}{\lambda_{\min}(\boldsymbol{C}_{n(11)})}\|\boldsymbol{\xi}_{S_0}\|_\infty \le (\lambda_n\sqrt{s_0}p_{\max}^{3/2})/\lambda_{\min}(\boldsymbol{C}_{n(11)})$$

using the ℓ_∞-ℓ_2 norm inequality. This proves (8.5) by Assumption 3.6(A). Next, we consider (8.6) and express (8.8) as $\boldsymbol{\xi}_{S_0^c} = \boldsymbol{W}_1 + \lambda_n^{-1}\boldsymbol{W}_2$, where

$$\boldsymbol{W}_1 = n^{-1}\boldsymbol{X}_{S_0^c}^{\mathrm{T}}\boldsymbol{X}_{S_0}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{\xi}_{S_0} \text{ and } \boldsymbol{W}_2 = n^{-1}\boldsymbol{X}_{S_0^c}^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{X}_{S_0^c}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{S_0^c}^{\mathrm{T}})\boldsymbol{\eta}.$$

For simplicity, we assume that the first s_0 groups are active and the remaining $(K-s_0)$ groups are noise. Then, supposing $\boldsymbol{W}_1 = (\boldsymbol{W}_{1,s_0+1}^{\mathrm{T}},\dots,\boldsymbol{W}_{1,K}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{W}_2 = (\boldsymbol{W}_{2,s_0+1}^{\mathrm{T}},\dots,\boldsymbol{W}_{2,K}^{\mathrm{T}})^{\mathrm{T}}$, and using $\|\boldsymbol{W}_{2,k}\| \le \sqrt{p_k}\|\boldsymbol{W}_{2,k}\|_\infty$, we can write for some $k \in S_0^c$, following from Assumption 3.7 and using Markov's inequality,

$$\begin{aligned}
\sup_{\sigma^*\in\mathcal{U}_n}\Pi\big(p_k^{-1/2}\|\boldsymbol{\xi}_k\| > 1\big|\boldsymbol{Y},\sigma^*\big) &\le \sup_{\sigma^*\in\mathcal{U}_n}\Pi\big(p_k^{-1/2}\|\boldsymbol{W}_{1,k}\| + \lambda_n^{-1}\|\boldsymbol{W}_{2,k}\|_\infty > 1\big|\boldsymbol{Y},\sigma^*\big)\\
&\le \sup_{\sigma^*\in\mathcal{U}_n}\Pi\big(\lambda_n^{-1}\|\boldsymbol{W}_{2,k}\|_\infty > 1-1+\nu\big|\boldsymbol{Y},\sigma^*\big)\\
&\le \frac{\sup_{\sigma^*\in\mathcal{U}_n}\mathbb{E}\big(\|\boldsymbol{W}_{2,k}\|_\infty\big|\boldsymbol{Y},\sigma^*\big)}{\lambda_n\nu}\\
&\le \frac{1}{\lambda_n\nu}\sup_{\sigma^*\in\mathcal{U}_n}\mathbb{E}\big[\|n^{-1}\boldsymbol{X}_k^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{X}_{S_0}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{X}_{S_0}^{\mathrm{T}})(\boldsymbol{\eta}-\boldsymbol{\mu})\|_\infty\big|\boldsymbol{Y},\sigma^*\big]\\
&\qquad + \|n^{-1}\boldsymbol{X}_k^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{X}_{S_0}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{X}_{S_0}^{\mathrm{T}})\boldsymbol{X}(\hat{\boldsymbol{\beta}}^{\mathrm{R}} - \boldsymbol{\beta}^0)\|_\infty\\
&= \mathcal{O}\big(\sqrt{\log(p-p_0)/n\lambda_n^2}\big),
\end{aligned}$$

where the first part of the inequality follows, by Assumption 3.3, as

$$\begin{aligned}
\sup_{\sigma^*\in\mathcal{U}_n} \operatorname{var}(\boldsymbol{W}_{2,k}|\boldsymbol{Y}) &= \operatorname{var}\big(n^{-1}\boldsymbol{X}_k^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{X}_{S_0}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{X}_{S_0}^{\mathrm{T}})\boldsymbol{\eta}\big|\boldsymbol{Y},\sigma^*\big) \\
&= \sigma^{*2}n^{-2}\boldsymbol{X}_k^{\mathrm{T}}(\boldsymbol{I}_n - \boldsymbol{X}_{S_0}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{X}_{S_0}^{\mathrm{T}})\boldsymbol{H}(a_n)(\boldsymbol{I}_n - \boldsymbol{X}_{S_0}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{X}_{S_0}^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{X}_k \\
&\leq \sigma^{*2}n^{-2}\|\boldsymbol{X}_k^{\mathrm{T}}\boldsymbol{H}(a_n)\boldsymbol{X}_k\| \\
&\leq \sigma^{*2}n^{-2}\|\boldsymbol{X}_k\|^2.
\end{aligned}$$

Thus, for any $k \in S_0^{\mathrm{c}}$, $\sup\{\Pi\big(p_k^{-1/2}\|\boldsymbol{\xi}_k\| > 1\big|\boldsymbol{Y},\sigma^*\big) : \sigma^* \in \mathcal{U}_n\} \to 0$ since $\log(p - p_0)/n\lambda_n^2 \to 0$ as $n \to \infty$. □

Proof of Lemma 3. Define the event $R_1 = \{\|\hat{\boldsymbol{\beta}}_k\| \geq \psi\|\boldsymbol{\beta}_k^0\| \text{ for } k \in S_0\}$. From Wei and Huang [2010], we know $P(\min_{k\in S_0}\|\hat{\boldsymbol{\beta}}_k\| \geq \psi\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|) \to 1$ as $n \to \infty$, where the inverse of $\|\hat{\boldsymbol{\beta}}_k\|$ is used as the weight in the group adaptive LASSO projection map, and hence needs to be controlled. Thus,

$$\begin{aligned}
&\sup_{\sigma^*\in\mathcal{U}_n} \Pi(A^{\mathrm{c}}|\boldsymbol{Y},\sigma^*) \\
&\leq \sup_{\sigma^*\in\mathcal{U}_n} \Pi(A^{\mathrm{c}}\cap R_1|\boldsymbol{Y},\sigma^*) + P(R_1^{\mathrm{c}}) \\
&\leq \sup_{\sigma^*\in\mathcal{U}_n} \Pi\big(\big\{\|(\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{Z}_{n(1)})_k\| + \sqrt{n}\|(\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{M}_{S_0})_k\| > \sqrt{n}\|\boldsymbol{\beta}_k^0\| \quad \text{for } k \in S_0\big\} \\
&\qquad\qquad \cap\{\|\hat{\boldsymbol{\beta}}_k\| \geq \psi\|\boldsymbol{\beta}_k^0\| \quad \text{for } k\in S_0\}\big|\boldsymbol{Y},\sigma^*\big) + o(1) \\
&\leq \sup_{\sigma^*\in\mathcal{U}_n} \Pi\big(\big\{\max_{k\in S_0}\|(\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{Z}_{n(1)})_k\| > \min_{k\in S_0}\big(\sqrt{n}\|\boldsymbol{\beta}_k^0\| - \sqrt{n}\|(\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{M}_{S_0})_k\|\big)\big\} \\
&\qquad\qquad \cap\{\|\hat{\boldsymbol{\beta}}_k\| \geq \psi\|\boldsymbol{\beta}_k^0\| \quad \text{for } k\in S_0\}\big|\boldsymbol{Y},\sigma^*\big) + o(1) \\
&\leq \sup_{\sigma^*\in\mathcal{U}_n} \Pi\Big(\Big\{\frac{1}{\sqrt{n}}\max_{k\in S_0}\|(\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{(1)}^{\mathrm{T}}\boldsymbol{\eta})_k\| > \sqrt{n}\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\| - \frac{s_0\lambda_n\sqrt{n}\lambda_{\min}^{-1}(\boldsymbol{C}_{n(11)})p_{\max}^{3/2}}{\psi\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|}\Big\}\Big|\boldsymbol{Y},\sigma^*\Big) + o(1),
\end{aligned}$$

since

$$\begin{aligned}
\min_{k\in S_0}\big(\sqrt{n}\|\boldsymbol{\beta}_k^0\| - \sqrt{n}\|(\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{M}_{S_0})_k\|\big) &\geq \sqrt{n}\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\| - \sqrt{n}\max_{k\in S_0}\|(\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{M}_{S_0})_k\| \\
&\geq \sqrt{n}\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\| - \frac{\sqrt{n}p_0}{\lambda_{\min}(\boldsymbol{C}_{n(11)})}\max_{k\in S_0}\|\boldsymbol{M}_k\| \\
&\geq \sqrt{n}\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\| - \frac{\sqrt{np_{\max}}\lambda_n s_0 p_{\max}}{\lambda_{\min}(\boldsymbol{C}_{n(11)})\psi\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|} \\
&= v_1, \text{ say.}
\end{aligned}$$

Under Assumption 3.6(C), we can bound v_1 as follows:

$$\begin{aligned}
v_1 &= \sqrt{n}\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\| - \frac{\sqrt{n}s_0\lambda_n(p_{\max})^{3/2}}{\lambda_{\min}(\boldsymbol{C}_{n(11)})\psi\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|} \\
&= \sqrt{n}\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\| + o(\sqrt{n}\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|).
\end{aligned}$$

Then, using Markov's inequality followed by a maximal inequality, we will have

$$\begin{aligned}
\sup_{\sigma^*\in\mathcal{U}_n}\Pi(A^{\mathrm{c}}|\boldsymbol{Y},\sigma^*) &\leq \frac{1}{v_1}\sup_{\sigma^*\in\mathcal{U}_n}\mathbb{E}\big(n^{-1/2}\max_{k\in S_0}\|(\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{(1)}^{\mathrm{T}}\boldsymbol{\eta})_k\|\big|\boldsymbol{Y},\sigma^*\big) \\
&\leq \frac{1}{v_1\sqrt{n}}\Big\{\sup_{\sigma^*\in\mathcal{U}_n}\mathbb{E}\big(\max_{k\in S_0}\|\big(\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{(1)}^{\mathrm{T}}(\boldsymbol{\eta}-\boldsymbol{\mu})\big)_k\|\big|\boldsymbol{Y},\sigma^*\big) + \max_{k\in S_0}\|\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{X}_{(1)}^{\mathrm{T}}\boldsymbol{\mu}\|\Big\} \\
&\leq K\frac{\sqrt{p_{\max}\log s_0}}{v_1\sqrt{n}} \to 0,
\end{aligned}$$

by Assumption 3.6(C). □

Proof of Lemma 4. We first define the event $R_2 = \{\|\hat{\boldsymbol{\beta}}_k\|^{-1} > r_n\}$, where r_n is the rate of convergence of the group LASSO estimator $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^{\mathrm{T}},\dots,\hat{\boldsymbol{\beta}}_K^{\mathrm{T}})^{\mathrm{T}}$. Then,

$$\begin{aligned}
&\sup_{\sigma^*\in\mathcal{U}_n}\Pi(B^{\mathrm{c}}|\boldsymbol{Y},\sigma^*) \\
&\leq \sup_{\sigma^*\in\mathcal{U}_n}\Pi(B^{\mathrm{c}}\cap R_2|\boldsymbol{Y},\sigma^*) + P(R_2^{\mathrm{c}}) \\
&\leq \sup_{\sigma^*\in\mathcal{U}_n}\Pi\Big(\Big\{\frac{1}{\sqrt{n}}\|(\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}(\boldsymbol{I}_n-\boldsymbol{P}_{S_0})\boldsymbol{\eta})_k\| \geq \lambda_n r_n\sqrt{np_k}\|\boldsymbol{1}_{p_k}\| - \sqrt{n}\|\boldsymbol{C}_{n(21)}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{M}_{S_0}\|\Big\}\Big|\boldsymbol{Y},\sigma^*\Big) + o(1) \\
&\leq \sup_{\sigma^*\in\mathcal{U}_n}\Pi\Big(\Big\{\frac{1}{\sqrt{n}}\|(\boldsymbol{X}_{S_0^{\mathrm{c}}}^{\mathrm{T}}(\boldsymbol{I}_n-\boldsymbol{P}_{S_0})\boldsymbol{\eta})_k\| \geq v_2\Big\}\Big|\boldsymbol{Y},\sigma^*\Big) + o(1),
\end{aligned}$$

where

$$
\begin{aligned}
v_2 =&\lambda_n r_n \sqrt{np_k}\|\mathbf{1}_{p_k}\| - \sqrt{n}\|\boldsymbol{C}_{n(21)}\boldsymbol{C}_{n(11)}^{-1}\boldsymbol{M}_{S_0}\| \\
\geq&\sqrt{n}\lambda_n r_n p_{\min} - \frac{\sqrt{n}p_0^2\lambda_n\sqrt{p_{\max}}}{\lambda_{\min}(\boldsymbol{C}_{n(11)})\psi\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|} \\
\geq&\sqrt{n}\lambda_n r_n p_{\min} - \frac{\sqrt{n}\lambda_n s_0^2(p_{\max})^{5/2}}{\lambda_{\min}(\boldsymbol{C}_{n(11)})\psi\min_{k\in S_0}\|\boldsymbol{\beta}_k^0\|} \\
=&\sqrt{n}\lambda_n r_n p_{\min}\big(1+o(1)\big)
\end{aligned}
$$

by Assumption 3.6(C).

Finally, using Markov's inequality and recalling that $p_{\max}\log(K-s_0)/(\sqrt{n}\lambda_n r_n p_{\min}) \to 0$ as $n\to\infty$, we can conclude that $\sup\{\Pi(B^{\mathrm{c}}|\boldsymbol{Y},\sigma^*) : \sigma^*\in\mathcal{U}_n\}\to 0$. □

Supplement C: Sparse Projection-posterior based on Group MCP penalty

Our existing proof for the group LASSO estimator already provides the appropriate scaffold: the argument proceeds through the basic inequality, stochastic score control (Lemma 6), the cone condition induced by the penalty, the compatibility condition, and the resulting estimation and prediction rates. To extend the argument to the *group MCP* estimator, we only need to modify the steps involving the penalty algebra. All other ingredients, including Lemma 6 and Lemma 5, remain unchanged.

For a group k with Euclidean norm $t=\|\boldsymbol{\beta}_k\|_2$, the group MCP penalty with parameters (λ,γ), where $\gamma>1$, is defined as

$$
P_{\lambda,\gamma}(t) = \begin{cases} \lambda t - \dfrac{t^2}{2\gamma}, & 0\leq t\leq \gamma\lambda, \\ \dfrac{1}{2}\gamma\lambda^2, & t>\gamma\lambda. \end{cases}
$$

Its group-level derivative is

$$P'_{\lambda,\gamma}(t) = \begin{cases} \lambda - \dfrac{t}{\gamma}, & 0 \le t \le \gamma\lambda, \\ 0, & t > \gamma\lambda. \end{cases}$$

Two key properties will be repeatedly used:

- For any $u, v \in \mathbb{R}^{p_k}$ with $t_u = \|u\|_2$ and $t_v = \|v\|_2$,

$$P_{\lambda,\gamma}(t_u) - P_{\lambda,\gamma}(t_v) \le \lambda(t_u - t_v) - \frac{1}{2\gamma}(t_u - t_v)_+^2 \le P_{\lambda,\gamma}(t_u) - P_{\lambda,\gamma}(t_v) \le \lambda(t_u - t_v);$$

- if $t_u, t_v \ge \gamma\lambda$, then

$$P_{\lambda,\gamma}(t_u) = P_{\lambda,\gamma}(t_v) = \tfrac{1}{2}\gamma\lambda^2,$$

  so in this region, the penalty exerts no shrinkage.

These properties imply a locally stronger curvature control than the group LASSO and complete bias elimination for sufficiently large groups. The fundamental basic inequality remains identical to the group LASSO case:

$$\frac{1}{n}\|\boldsymbol{X}\boldsymbol{\beta}^0 - \boldsymbol{X}\boldsymbol{\beta}^*\|_2^2 \le \frac{2}{n}\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{\beta}^0 - \boldsymbol{\beta}^*) + \sum_{k=1}^{K}\left\{P_{\lambda,\gamma}(\|\boldsymbol{\beta}_k^0\|_2) - P_{\lambda,\gamma}(\|\boldsymbol{\beta}_k^*\|_2)\right\}.$$

The stochastic term is controlled by Lemma 6 with the same choice of $\lambda \ge \lambda_0(x)$. The only modification concerns the penalty difference. Where the group LASSO argument uses $\sum_{k=1}^{K}\lambda\sqrt{p_k}\big(\|\boldsymbol{\beta}_k^0\|_2 - \|\boldsymbol{\beta}_k^*\|_2\big)$, we replace it with the group MCP bound

$$P_{\lambda,\gamma}(\|\boldsymbol{\beta}_k^0\|_2) - P_{\lambda,\gamma}(\|\boldsymbol{\beta}_k^*\|_2) \le \lambda\big(\|\boldsymbol{\beta}_k^0\|_2 - \|\boldsymbol{\beta}_k^*\|_2\big) - \frac{1}{2\gamma}\big(\|\boldsymbol{\beta}_k^*\|_2 - \|\boldsymbol{\beta}_k^0\|_2\big)_+^2.$$

Summing over all groups yields

$$\sum_{k=1}^{K}\Big\{P_{\lambda,\gamma}(\|\boldsymbol{\beta}_k^0\|_2) - P_{\lambda,\gamma}(\|\boldsymbol{\beta}_k^*\|_2)\Big\} \le \lambda\sum_{k=1}^{K}\sqrt{p_k}\big(\|\boldsymbol{\beta}_k^0\|_2 - \|\boldsymbol{\beta}_k^*\|_2\big) - \frac{1}{2\gamma}\sum_{k=1}^{K}\big(\|\boldsymbol{\beta}_k^*\|_2 - \|\boldsymbol{\beta}_k^0\|_2\big)_+^2.$$

This preserves the decomposable structure used to establish the cone condition, but with an additional nonnegative curvature term that sharpens the inequality and reduces estimation bias. Following the same steps as in the group LASSO case, we obtain

$$\frac{1}{n}\|\boldsymbol{X}\Delta\|_2^2 + \lambda \sum_{k \notin S_0} \sqrt{p_k}\|\boldsymbol{\beta}_k^*\|_2 + \frac{1}{2\gamma}\sum_{k=1}^{K}\left(\|\boldsymbol{\beta}_k^*\|_2 - \|\boldsymbol{\beta}_k^0\|_2\right)_+^2 \le \frac{3\lambda}{2}\sum_{k \in S_0} \sqrt{p_k}\|\Delta_k\|_2,$$

where $\Delta_k = \boldsymbol{\beta}_k^* - \boldsymbol{\beta}_k^0$. Dropping the nonnegative curvature term yields the usual inequality $\sum_{k \notin S_0} \sqrt{p_k}\|\boldsymbol{\beta}_k^*\|_2 \le (3/2)\sum_{k \in S_0} \sqrt{p_k}\|\Delta_k\|_2$, while retaining it leads to a strictly tighter constant. Under Assumption 3.4, we have

$$\sum_{k \in S_0} \sqrt{p_k}\|\Delta_k\|_2 \le \frac{\sqrt{\sum_{k \in S_0} p_k}}{\phi(S_0)} \frac{\|\boldsymbol{X}\Delta\|_2}{\sqrt{n}}.$$

Substituting into the basic inequality and applying $2ab \le a^2 + b^2$ yields

$$\frac{1}{2n}\|\boldsymbol{X}\Delta\|_2^2 + \frac{\lambda}{2}\sum_{k=1}^{K} \sqrt{p_k}\|\Delta_k\|_2 \lesssim \frac{\lambda^2}{\phi^2(S_0)} \sum_{k \in S_0} p_k.$$

Hence, the estimation and prediction rates match those obtained under the group LASSO penalty, with improved constants due to the concavity of the MCP penalty.

For exact support recovery, we impose a groupwise β-min condition:

$$\min_{k \in S_0} \|\boldsymbol{\beta}_k^0\|_2 \ge (1+\delta)\gamma\lambda,$$

together with the weighted group irrepresentability condition. This ensures that the active groups lie within the flat region of the MCP penalty (no shrinkage), while inactive groups are shrunk to zero via the KKT conditions. This is the MCP analogue of Assumption 3.6(B) used for SCAD.

Thus, the overall structure of the proof remains unchanged, and the group MCP yields identical convergence rates with reduced estimation bias.

Supplement D: Debiased Projection-posterior for group SCAD and adaptive group LASSO

Assumption 8.1 (GS and AGL conditions)**.** *Assume the design and regularity conditions used in Theorem 3 hold. In addition:*

(GS) *Let* $\hat{\boldsymbol{\beta}}^{\mathrm{GS}}$ *be any stationary point of the group SCAD objective. Suppose there is a beta-min constant* $c_0 > 0$ *such that* $\min_{k \in S_0} \|\beta_k^0\| \geq c_0$*, and* $\hat{\boldsymbol{\beta}}^{\mathrm{GS}}$ *is selection consistent and* $\|\hat{\boldsymbol{\beta}}^{\mathrm{GS}} - \boldsymbol{\beta}^0\|_2 = \mathcal{O}_P(\lambda_n \sum_{k=1}^{s_0} p_k / \sqrt{p_{\min}})$ *with* λ_n *as defined in Theorem 1(2).*

(AGL) *Let* $\hat{\boldsymbol{\beta}}^{\mathrm{AGL}}$ *be the adaptive group LASSO estimator with weights* $w_k = \|\tilde{\beta}_k\|^{-\gamma}$ $(\gamma > 0)$*. Assume the initial estimator* $\tilde{\beta}$ *is zero-consistent: for some rate* $r_n \to \infty$, $\max_{k \in S_0^c} \|\tilde{\beta}_k\| = \mathcal{O}_P(r_n^{-1})$ *and* $\min_{k \in S_0} \|\tilde{\beta}_k\| \geq c_0$ *w.h.p., so that* $w_k = \mathcal{O}_P(r_n^{\gamma})$ *for* $k \in S_0^c$ *and* $w_k = \mathcal{O}_P(1)$ *for* $k \in S_0$*. With tuning satisfying the condition in Theorem 1(3), the estimator is selection consistent and* $\|\hat{\boldsymbol{\beta}}^{\mathrm{AGL}} - \boldsymbol{\beta}^0\|_2 = \mathcal{O}_P(\lambda_n \sum_{k=1}^{s_0} p_k / \sqrt{p_{\min}})$.

Define the remainder terms

$$\boldsymbol{\Delta}^{\mathrm{GS}} = (\hat{\boldsymbol{\Theta}}\hat{\boldsymbol{\Sigma}} - I_p)(\hat{\boldsymbol{\beta}}^{\mathrm{GS}} - \boldsymbol{\beta}^0),\ \boldsymbol{\Delta}^{\mathrm{AGL}} = (\hat{\boldsymbol{\Theta}}\hat{\boldsymbol{\Sigma}} - I_p)(\hat{\boldsymbol{\beta}}^{\mathrm{AGL}} - \boldsymbol{\beta}^0),$$

and the corresponding debiased projections

$$\boldsymbol{\beta}^{**}_{\mathrm{DGS}} = \boldsymbol{\beta}^0 + \frac{1}{n}\hat{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\eta} - \boldsymbol{\Delta}^{\mathrm{GS}},\ \boldsymbol{\beta}^{**}_{\mathrm{DAGL}} = \boldsymbol{\beta}^0 + \frac{1}{n}\hat{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\eta} - \boldsymbol{\Delta}^{\mathrm{AGL}}.$$

Theorem 5 (BvM for debiased GS/AGL projection-posteriors)**.** *Suppose the assumptions of Theorem 3 on the construction of* $\hat{\boldsymbol{\Theta}}$ *and the posterior covariance* $\boldsymbol{H}(a_n)$ *hold, and in addition Assumption 8.1 holds.*

(i) **Group SCAD.** *We have* $\|\boldsymbol{\Delta}^{\mathrm{GS}}\|_\infty = o_P(n^{-1/2})$ *and*

$$\sup_B \left| \Pi\big(\sigma^{-1}\sqrt{n}\,(\boldsymbol{\beta}^{**}_{\mathrm{DGS}} - \boldsymbol{\beta}^0) \in B \,\big|\, \boldsymbol{Y}\big) - \mathcal{N}_p\big(B; \boldsymbol{m}, \boldsymbol{V}\big) \right| \ \to\ 0,$$

where $\boldsymbol{m} = n^{-1/2}\hat{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\mu}$ *and* $\boldsymbol{V} = n^{-1}\hat{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{H}(a_n)\hat{\boldsymbol{\Theta}}$ *are same as in Theorem 3.*

(ii) **Adaptive group LASSO.** *We have* $\|\boldsymbol{\Delta}^{\mathrm{AGL}}\|_\infty = o_P(n^{-1/2})$ *and*

$$\sup_B \left| \Pi\big(\sigma^{-1}\sqrt{n}\,(\boldsymbol{\beta}^{**}_{\mathrm{DAGL}} - \boldsymbol{\beta}^0) \in B \,\big|\, \boldsymbol{Y}\big) - \mathcal{N}_p\big(B;\, \boldsymbol{m}, \boldsymbol{V}\big) \right| \;\to\; 0,$$

with the same Gaussian limit $\mathcal{N}_p(\cdot; \boldsymbol{m}, \boldsymbol{V})$ *as in part* (i) *and Theorem 3.*

Corollary 2 (Coverage)**.** *Under the conditions of Theorem 5, component-wise credible intervals and ellipsoidal credible sets obtained from the GS or AGL debiased projection-posteriors have asymptotically exact frequentist coverage.*

The proofs are identical to those of Theorem 3 and Corollary 1 and hence, we skip them to avoid repetition.