Supplementary Material for "KNOWLEDGE TRANSFER FOR SPARSE PARTIAL LINEAR MODELS WITH PRIVACY GUARANTEE : ESTIMATION, INFERENCE AND SIMULTANEOUS TESTING"

This supplementary material is organized as follows. We first present the algorithmic components of the proposed procedures, then describe the privatized BIC methods for sparsity selection, report an additional simulation study, and finally provide the technical details supporting the theoretical results.

A1. Algorithms

This section provides the basic algorithmic components used throughout the proposed private transfer procedures.

Algorithm 1 Hard-thresholding ($\mathrm{HT}(\boldsymbol{v}, s)$)

INPUT: Vector-valued function $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{X}) \in \mathbb{R}^p$ with data $\boldsymbol{X}$, sparsity s.
for $j = 1, \ldots, p$ **do**

$$\boldsymbol{v}_j^{t+1} = \begin{cases} \boldsymbol{v}_j^{t+0.5} & \text{if } |\boldsymbol{v}_j^{t+0.5}| \text{ is among the } s \text{ largest values of } \{|\boldsymbol{v}_j^{t+0.5}|\}_{j\in[p]}, \\ 0 & \text{otherwise.} \end{cases}$$

end for
OUTPUT: $\boldsymbol{v}$

Algorithm 2 Noisy Hard-thresholding (NoisyHT $(\boldsymbol{v}, s, \lambda, \epsilon, \delta)$)

INPUT: Vector-valued function $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{X}) \in \mathbb{R}^p$ with data $\boldsymbol{X}$, sparsity s, privacy parameters (ϵ, δ), sensitivity λ.
Initialization: Set $\mathcal{S} = \emptyset$.
for $i = 1, \ldots, s$ **do**
  Generate noise $w_{i1}, w_{i2}, \ldots, w_{ip} \overset{\text{i.i.d.}}{\sim} \text{Laplace}(\lambda \cdot \frac{2\sqrt{3s\log(1/\delta)}}{\epsilon})$.
  Append $j^* = \text{argmax}_{j \in [p] \setminus \mathcal{S}}(|v_j| + w_{ij})$ to $\mathcal{S}$.
end for
Generate noise $\tilde{w}_1, \ldots, \tilde{w}_p \overset{\text{i.i.d.}}{\sim} \text{Laplace}(\lambda \cdot \frac{2\sqrt{3s\log(1/\delta)}}{\epsilon})$.
for $j = 1, \ldots, p$ **do**

$$\tilde{v}_j^{t+1} = \begin{cases} v_j + \tilde{w}_j & \text{if } j \in \mathcal{S}, \\ 0 & \text{otherwise.} \end{cases}$$

end for
OUTPUT: $\tilde{\boldsymbol{v}}$

A2. Differentially Private Sparsity Selection via Privatized BIC

This section describes the privatized BIC criteria used to select the sparsity levels for both the transfer estimator and the asymptotic variance estimator.

Sparsity of $\boldsymbol{\beta}^{(0)}$

For sparsity selection, we construct a dyadic grid with a chosen maximum grid index M_s,

$$s_a = 2^a, \qquad a = 0, \ldots, M_s,$$

and perform transfer estimation independently for each candidate sparsity level. After obtaining the sequence of transfer estimators $\{\hat{\boldsymbol{\beta}}(a) : 0 \leq a \leq M_s\}$ from the differentially private transfer learning procedure, we perform sparsity selection using a privatized Bayesian Information Criterion (PBIC) computed solely on the target sample. This step is designed to (i) preserve differential privacy under the Gaussian mechanism, (ii) control the sensitivity of the goodness-of-fit term via truncation, and (iii) account explicitly for privacy-induced variance inflation in the model complexity penalty.

Let $\{(\boldsymbol{X}_i^{(0)}, \breve{Y}_i^{(0)})\}_{i=1}^{n_0}$ denote the target sample. For each candidate sparsity level a, define the truncated residual sum of squares:

$$\text{RSS}(\hat{\boldsymbol{\beta}}(a)) = \sum_{i=1}^{n_0} \left(\Pi_{R_Y}(\breve{Y}_i^{(0)}) - \Pi_{R_0}((\boldsymbol{X}_i^{(0)})^\top \hat{\boldsymbol{\beta}}(a)) \right)^2.$$

This bounded sensitivity $\mathrm{RSS}(\hat{\boldsymbol{\beta}}(a))$ allows us to apply the Gaussian mechanism. We therefore release the privatized goodness-of-fit $\widetilde{\mathrm{RSS}}(\hat{\boldsymbol{\beta}}(a)) = \mathrm{RSS}(\hat{\boldsymbol{\beta}}(a)) + \Delta_C,\ \Delta_C \sim \mathrm{N}\left(0, \frac{2\log(3.75/\delta)\,(R_Y+R_0)^4}{(\epsilon/3)^2}\right)$. Therefore, by the Gaussian mechanism with privacy budget $(\epsilon/3, \delta/3)$, the above privatized goodness-of-fit is $(\epsilon/3, \delta/3)$-differentially private for each fixed a.

Let s_a denote the sparsity associated with $\hat{\boldsymbol{\beta}}(a)$. The privatized BIC criterion of $\hat{\boldsymbol{\beta}}(a)$ is defined as:

$$\mathrm{PBIC}(\hat{\boldsymbol{\beta}}(a)) = \widetilde{\mathrm{RSS}}(\hat{\boldsymbol{\beta}}(a)) + C\left(\log(p)\log(n_0)s_a + \frac{\log^2(p)s_a^2\log(3.75/\delta)\log^7(n_0)}{n_0(\epsilon/3)^2}\right).$$

The penalty consists of two components: $\log(p)\log(n_0)s_a$ is the high-dimensional complexity term that controls overfitting under ultra-high dimensional scaling. $\log^2(p)s_a^2\log(3.75/\delta)\log^7(n_0)/n_0(\epsilon/3)^2$ is a privacy-adjustment term. It compensates for the additional stochastic fluctuation introduced by the Gaussian noise and the privatized transfer estimation, ensuring that the selection procedure remains consistent under privacy constraints. The constant $C > 0$ is chosen as the way in Theorem 1. The selected sparsity level is

$$\hat{a} = \arg\min_{0\le a\le M_s} \mathrm{PBIC}(\hat{\boldsymbol{\beta}}(a)), \qquad \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{a}).$$

Thus, model selection is conducted entirely on the target sample using a Gaussian-privatized, truncation-stabilized information criterion. The complete procedure is concluded in Algorithm 3.

Sparsity of $\boldsymbol{\Theta}_j^{(0)}$

For every given $j \in [p]$, we construct a sequence of candidate estimators $\{\hat{\boldsymbol{\Theta}}_j(b) : 0 \le b \le M_\Theta\}$ using the same differentially private iterative hard-thresholding scheme as in the transfer step, but with the sparsity level fixed at $s_{j,b} = 2^b$. For each candidate sparsity level, model selection is carried out via a privatized BIC criterion. Specifically, we compute a truncated quadratic loss measuring the fidelity of the linear projection of $\check{\boldsymbol{X}}_{\cdot,j}^{(0)}$ onto $\check{\boldsymbol{X}}^{(0)}$, and add Gaussian noise calibrated to its ℓ_2-sensitivity under truncation radius $\tilde{R}$. The resulting privatized goodness-of-fit is combined with a high-dimensional complexity penalty

$$\log(p)\log(n_0)s_{j,b}$$

and a privacy-adjustment term of order

$$\frac{\log^2(p)s_{j,b}^2\log(1/\delta)\log^7(n_0)}{n_0^2\epsilon^2},$$

Algorithm 3 Transfer High-dimensional Partial Linear Regression with Differential Privacy and BIC-based Sparsity Selection

INPUT: $\{(\boldsymbol{X}_i^{(0)}, \boldsymbol{W}_i^{(0)}, Y_i^{(0)})\}$, $\{(\boldsymbol{X}_i^{(k)}, Y_i^{(k)}), k \in [K]\}$, T, ρ_1, (ϵ, δ), $\boldsymbol{\beta}^0$, M_s, $\mathcal{A}$, L, R_Y, R_0, R_k, R_d.

Define sparsity candidates $s_a = 2^a$, $a = 0, \ldots, M_s$.

Residualization:

$\hat{m}^{(init)} = \arg\min_{m \in \mathcal{F}} \frac{1}{2n_0} \sum_{i=1}^{n_0} (Y_i^{(0)} - m(\boldsymbol{W}_i^{(0)}))^2 + \frac{\lambda_0}{2} \|m\|_{\mathcal{F}}^2$.

$\breve{\boldsymbol{Y}}^{(0)} = \Pi_{R_Y}\left(\boldsymbol{Y}^{(0)} - \hat{\boldsymbol{m}}^{(init)}(\boldsymbol{W}^{(0)})\right) + \boldsymbol{\Delta}_Y,\ \boldsymbol{\Delta}_Y \sim \mathrm{N}\left(\boldsymbol{0}, \frac{8\log(3.75/\delta) R_Y^2}{(\epsilon/3)^2} \mathbf{I}_{n_0}\right)$.

for $k' \in \mathcal{A}$ **do**

  $\breve{Y}_i^{(k')} = Y_i^{(k')}$.

end for

for $a = 0, \ldots, M_s$ **do**

  Initialize $\boldsymbol{\beta}^0$.

  Split each dataset into T blocks $\mathcal{I}^{(k')[t]}$.

  for $t = 0, \ldots, T-1$ **do**

    for $k' \in \{0\} \cup \mathcal{A}$ **do**

      $\tilde{\boldsymbol{G}}_{k'}^t = \frac{n_{k'}}{N}\left(\frac{1}{\lfloor n_{k'}/T \rfloor} \sum_{i \in \mathcal{I}^{(k')[t]}} \Pi_{R_{k'}}((\boldsymbol{X}_i^{(k')})^\top \boldsymbol{\beta}^t - \breve{Y}_i^{(k')}) \Pi_{R_d}(\boldsymbol{X}_i^{(k')}) + \boldsymbol{\Delta}_t^{(k')}\right)$, $\boldsymbol{\Delta}_t^{(k')} \sim \mathrm{N}\left(\boldsymbol{0}, \frac{8\log(3.75/\delta) T^2 R_d^2 R_{k'}^2}{n_{k'}^2 (\epsilon/3)^2} \mathbf{I}_p\right)$.

    end for

    $\boldsymbol{\beta}^{t+0.5} = \boldsymbol{\beta}^t - \rho_1 \sum_{k' \in \{0\} \cup \mathcal{A}} \tilde{\boldsymbol{G}}_{k'}^t$.

    $\boldsymbol{\beta}^{t+1} = \mathrm{HT}(\boldsymbol{\beta}^{t+0.5}, s_a)$.

  end for

  $\hat{\boldsymbol{\beta}}(a) = \boldsymbol{\beta}^T$.

  Compute truncated RSS on target: $\mathrm{RSS}(\hat{\boldsymbol{\beta}}(a)) = \sum_{i=1}^{n_0}\left(\Pi_{R_Y}(\breve{Y}_i^{(0)}) - \Pi_{R_0}((\boldsymbol{X}_i^{(0)})^\top \hat{\boldsymbol{\beta}}(a))\right)^2$.

  $\mathrm{PBIC}(\hat{\boldsymbol{\beta}}(a)) = \mathrm{RSS}(\hat{\boldsymbol{\beta}}(a)) + C\left(\log(p)\log(n_0)s_a + \frac{\log^2(p) s_a^2 \log(3.75/\delta) \log^7(n_0)}{n_0 (\epsilon/3)^2}\right) + \Delta_C,\ \Delta_C \sim \mathrm{N}\left(0, \frac{2\log(3.75/\delta)(R_Y + R_0)^4}{(\epsilon/3)^2}\right)$.

end for

$\hat{a} = \arg\min_{0 \le a \le M_s} \mathrm{PBIC}(\hat{\boldsymbol{\beta}}(a))$.

OUTPUT: $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{a})$.

which compensates for stochastic inflation induced by the Gaussian mechanism. The privatized BIC criterion for $\hat{\mathbf{\Theta}}_j(b)$ is defined as

$$\begin{aligned}\mathrm{PBIC}_j(\hat{\mathbf{\Theta}}_j(b)) = n_0 \left[\frac{1}{2n_0} \sum_{i=1}^{n_0} (\hat{\mathbf{\Theta}}_j(b))^\top \check{\boldsymbol{X}}_i^{(0)} (\check{\boldsymbol{X}}_i^{(0)})^\top \hat{\mathbf{\Theta}}_j(b) - (\hat{\mathbf{\Theta}}_j(b))^\top \mathbf{e}_j \right] \\ + C\left(\log(p)\log(n_0) s_{j,b} + \frac{\log^2(p) s_{j,b}^2 \log(2.5/\delta) \log^7(n_0)}{n_0^2(\epsilon/2)^2} \right) + \Delta_{\Sigma,b},\end{aligned}$$

where $\Delta_{\Sigma,b} \sim \mathrm{N}\left(0, \frac{2\tilde{R}^4 \log(2.5/\delta)}{(\epsilon/2)^2}\right)$. The selected sparsity level $\hat{b}_j$ minimizes this privatized criterion, and the corresponding estimator is $\hat{\mathbf{\Theta}}_j = \hat{\mathbf{\Theta}}_j(\hat{b}_j)$. Combined with the privatized residual variance estimator, this yields the final differentially private asymptotic variance estimate

$$\hat{V}_j^{(dp)} = \left(\hat{\Theta}_{j,j} \hat{\sigma}^2\right)^{1/2}.$$

This procedure ensures that sparsity adaptation for each matrix column is fully data-driven, privacy-preserving under the Gaussian mechanism, and theoretically aligned with the high-dimensional transfer learning framework. The complete procedure is concluded in Algorithm 4.

Algorithm 4 Asymptotic Variance Estimation with PBIC Selection

INPUT: Target sample $\{\check{\boldsymbol{X}}_i^{(0)}, \check{Y}_i^{(0)}\}_{i\in[n_0]}$, initial $\hat{\boldsymbol{\beta}}$, step size ρ_2, privacy parameters (ϵ, δ), noise scale B, iterations T, truncation level $\tilde{R}$, feasibility radius C, initial vector $\boldsymbol{\Theta}_j^0$.

Split $[n_0]$ into T disjoint subsets $\{\mathcal{I}^{(0)[t]}\}_{t=0}^{T-1}$. Set $M_\Theta = \lfloor \log_2(\sqrt{n_0}/\log^2 p)\rfloor$.

for $b = 0$ to M_Θ **do**

  $s_{j,b} = 2^b$, initialize $\boldsymbol{\Theta}_j^0(b) = \boldsymbol{\Theta}_j^0$.

  for $t = 0$ to $T-1$ **do**

    Gradient step:

$$\boldsymbol{\Theta}_j^{t+0.5}(b) = \boldsymbol{\Theta}_j^{t}(b) - \rho_2\left(\mathbf{e}_j - \frac{1}{|\mathcal{I}^{(0)[t]}|}\sum_{i\in\mathcal{I}^{(0)[t]}} \check{\boldsymbol{X}}_i^{(0)}\Pi_{\tilde{R}}\big((\check{\boldsymbol{X}}_i^{(0)})^\top\boldsymbol{\Theta}_j^t(b)\big)\right).$$

    Noisy hard-thresholding:

$$\boldsymbol{\Theta}_j^{t+1}(b) = \Pi_C\Big(\mathrm{NoisyHT}\left(\boldsymbol{\Theta}_j^{t+0.5}(b), s_{j,b}, \tfrac{\rho_2 TB}{n_0}, \tfrac{\epsilon}{2T(M_\Theta+1)}, \tfrac{\delta}{2T(M_\Theta+1)}\right)\Big).$$

  end for

  $\hat{\boldsymbol{\Theta}}_j(b) = \boldsymbol{\Theta}_j^T(b)$.

  Define

$$\begin{aligned}\mathrm{PBIC}_j(\hat{\boldsymbol{\Theta}}_j(b)) &= n_0\left[\frac{1}{2n_0}\sum_{i=1}^{n_0}(\hat{\boldsymbol{\Theta}}_j(b))^\top\check{\boldsymbol{X}}_i^{(0)}(\check{\boldsymbol{X}}_i^{(0)})^\top\hat{\boldsymbol{\Theta}}_j(b) - (\hat{\boldsymbol{\Theta}}_j(b))^\top\mathbf{e}_j\right] \\ &\quad + C\left(\log(p)\log(n_0)s_{j,b} + \frac{\log^2(p)s_{j,b}^2\log(2.5/\delta)\log^7(n_0)}{n_0^2(\epsilon/2)^2}\right) + \Delta_{\Sigma,b},\end{aligned}$$

  where $\Delta_{\Sigma,b}\sim \mathrm{N}\Big(0, \frac{2(M_\Theta+1)\tilde{R}^4\log(2.5/\delta)}{(\epsilon/2)^2}\Big)$.

end for

$\hat{b}_j = \arg\min_{0\le b\le M_\Theta}\mathrm{PBIC}_j(\hat{\boldsymbol{\Theta}}_j(b))$, $\hat{\boldsymbol{\Theta}}_j = \hat{\boldsymbol{\Theta}}_j(\hat{b}_j)$.

Compute privatized residual variance $\hat{\sigma}^2$.

OUTPUT: $\hat{V}_j^{(dp)} = \big(\hat{\Theta}_{j,j}\hat{\sigma}^2\big)^{1/2}$.

A3. Simulation Results for Configuration 4

This section exhibits an additional simulation setting that illustrates the role of transfer learning when the control variables and predictors are correlated.

- **Configuration 4**: Instead of letting $\boldsymbol{X}^{(0)} \perp \boldsymbol{W}^{(0)}$, we assume that the $X_1^{(0)} = |W_1^{(0)} + W_2^{(0)}|, X_2^{(0)} = |W_1^{(0)} - W_2^{(0)}|$. The remaining predictors are generated in the same manner as the first scenario, i.e., $\boldsymbol{X}_{3:p}^{(0)} \sim \mathrm{N}(\mathbf{0}, \Sigma_{3:p})$.

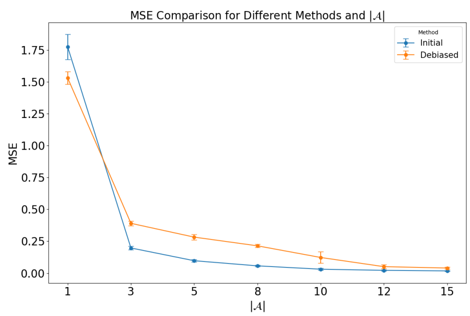

Figure 1: The necessity of transfer learning

In Configuration 4, we find that if the control variables are highly correlated with the predictors, the target-only estimation may not perform as effectively as it did in other configurations. In such case, transfer learning becomes essential for achieving acceptable results. Specifically, we present the MSE of Initial, Debiased and DP-Debiased. All estimators exhibit progressively better performance as more transferable sources are incorporated, compared to the target-only scenario (i.e., $|\mathcal{A}| = 1$).

A4. Technical Details

This section presents the auxiliary theoretical details, including technical lemmas, the derivation of the e-value, and all the proofs of the main theoretical results.

A4.1. Necessary Lemmas

We first provide the auxiliary lemmas used in the proofs of the main theoretical results, including nonparametric estimation bounds, hard-thresholding properties, and technical concentration inequalities.

Lemma 1. Let $\tilde{r}_n > 0$ be the smallest positive quantity satisfying the critical inequality

$$\mathcal{G}_n(\tilde{r}_n; \mathcal{F}) \leq \frac{\tilde{r}_n^2}{2C_\sigma},$$

where C_σ is the sub-Gaussian parameter for the residual term in corresponding regression. Then for any $t_n \geq \tilde{r}_n$, we have: (i) The estimator $\hat{f}_j$ satisfies that $\|\hat{f}_j(\boldsymbol{W}^{(0)}) - f_j(\boldsymbol{W}^{(0)})\|_n^2 = O_p(\tilde{s}_j d t_n^2)$, with $\tilde{s}_j$ defined in Assumption 4; (ii) The estimator $\hat{m}^{(init)}$ satisfies that $\|\hat{m}^{(init)}(\boldsymbol{W}^{(0)}) - (\boldsymbol{f}(\boldsymbol{W}^{(0)}))^\top \boldsymbol{\beta}^{(0)} - g(\boldsymbol{W}^{(0)})\|_n^2 = O_p(s_0 d t_n^2)$; (iii) Moreover, if $\|(\boldsymbol{X}^{(0)} - \boldsymbol{f}(\boldsymbol{W}^{(0)}))^\top \boldsymbol{\beta}^{(0)}\|_n^2 = O_p(s_0 d t_n^2)$, then $\|\hat{m}^{(init)}(\boldsymbol{W}^{(0)}) - (\boldsymbol{X}^{(0)})^\top \boldsymbol{\beta}^{(0)} - g(\boldsymbol{W}^{(0)})\|_n^2 = O_p(s_0 d t_n^2)$; (iv) The estimator $\hat{g}$ satisfies that $\|\hat{g}(\boldsymbol{W}^{(0)}) - g(\boldsymbol{W}^{(0)})\|_n^2 = O_p(s_0 d t_n^2)$. Here $\|\cdot\|_n$ represents the empirical loss which takes the average over $\boldsymbol{W}_i^{(0)}$, e.g. $\|\hat{g}(\boldsymbol{W}^{(0)}) - g(\boldsymbol{W}^{(0)})\|_n^2 = \sum_{i=1}^{n_0} (\hat{g}(\boldsymbol{W}_i^{(0)}) - g(\boldsymbol{W}_i^{(0)}))^2 / n_0$.

Proof. Recall that

$$\hat{f}_j = \arg\min_{f \in \mathcal{F}} [\frac{1}{2n_0} \sum_{i=1}^{n_0} \{X_{ij}^{(0)} - f(\boldsymbol{W}_i^{(0)})\}^2],\ j = 1, \dots, p,$$

and

$$\hat{m}^{(init)} = \arg\min_{m \in \mathcal{F}} [\frac{1}{2n_0} \sum_{i=1}^{n_0} \{Y_i^{(0)} - m(\boldsymbol{W}_i^{(0)})\}^2] + \frac{\lambda_0}{2} \|m\|_{\mathcal{F}}^2.$$

Lemma S.1, Lemma S.2 and Lemma S.10 in Zhu (2017) have shown the first two results in (i) and (ii), and the existence of t_n is given by Lemma A5 in Zhu et al. (2019). The triangle inequality gives (iii) under the stated additional condition. Combining (i) and (ii), and subsequently applying the triangle inequality, immediately leads to (iv). □

Lemma 2. Consider vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^p$, another vector $\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w}$, a set $S_1 \subseteq S'$ with S' as the indices of top-s' entries of $|\boldsymbol{u}|$ (absolute value for each entry), and another set $S_2 \subseteq (S')^C$ with $|S_1| = |S_2|$, where $s' \in \mathbb{N}_+$. Then for any $c \in (0, 1)$,

$$(1-c)\|\boldsymbol{v}_{S_2}\|_2^2 - \frac{1}{c}\|\boldsymbol{w}_{S_2}\|_2^2 \leq \|\boldsymbol{u}_{S_1}\|_2^2 \leq (1+c)\|\boldsymbol{v}_{S_1}\|_2^2 + \left(1 + \frac{1}{c}\right)\|\boldsymbol{w}_{S_1}\|_2^2,$$

which implies

$$\|\boldsymbol{v}_{S_2}\|_2^2 \leq \frac{1+c}{1-c}\|\boldsymbol{v}_{S_1}\|_2^2 + \frac{1}{1-c}\left[\frac{1}{c}|S_2| + \left(1 + \frac{1}{c}\right)|S_1|\right]\|\boldsymbol{w}\|_\infty^2.$$

Proof. This is the Lemma 20 in Li et al. (2024). □

A4.2. Proof of Theorem 1

Proof. Similar to Li et al. (2024), we prove our result from two aspects: differential privacy guarantee and algorithm consistency. For the first part, we construct a sequence of events and demonstrate that they occur with high probability, which shows the reasonable scale of the added noise. For the second part, we derive an upper bound for the estimation error. Before proceeding, we denote the sample splitting set as $\mathcal{D}^{[t]} = \{(\boldsymbol{X}_i^{(k)t}, Y_i^{(k)t})\}_{i\in\mathcal{I}^{(k)[t]}} := \{(\boldsymbol{X}^{(k)t}, \boldsymbol{Y}^{(k)t})\}$, and write $n_{k,t} = |\mathcal{I}^{(k)[t]}|$, so that $\sum_{t=0}^{T-1} n_{k,t} = n_k$.

(I) In the first part, we define the following four events, which will be used in the subsequent analysis. First, we need to upper and lower bound the eigenvalues of sample covariance matrix. The sample covariance matrix is defined as

$$\hat{\boldsymbol{\Sigma}}^{(k)[t]} = \frac{1}{n_{k,t}} \sum_{i\in\mathcal{I}^{(k)[t]}} \boldsymbol{X}_i^{(k)} (\boldsymbol{X}_i^{(k)})^\top.$$

Also denote that

$$\hat{\boldsymbol{\Sigma}}_{\mathcal{A}}^{[t]} = \sum_{k\in\{0\}\cup\mathcal{A}} \frac{n_k}{N} \hat{\boldsymbol{\Sigma}}^{(k)[t]}, \qquad \boldsymbol{\Sigma}_{\mathcal{A}} = \mathbb{E}(\hat{\boldsymbol{\Sigma}}_{\mathcal{A}}^{[t]}).$$

Let $\mathcal{E}_1$ be the event that the sample covariance matrix is well-conditioned and closely approximates the population covariance matrix, i.e.,

$$\begin{aligned}\mathcal{E}_1 =& \{\underline{\gamma} \le \lambda_{\min}(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}',\mathcal{S}'}^{(k)[t]}) \le \lambda_{\max}(\hat{\boldsymbol{\Sigma}}_{\mathcal{S}',\mathcal{S}'}^{(k)[t]}) \le \overline{\gamma}, \forall k \in \{0\}\cup\mathcal{A}, \forall \mathcal{S}' \subseteq [p] \text{ with } |\mathcal{S}'| \le 3s', \forall t = 0,\dots,T-1\} \\ & \bigcap \{\underline{\gamma} \le \lambda_{\min}((\hat{\boldsymbol{\Sigma}}_{\mathcal{A}})_{\mathcal{S}',\mathcal{S}'}^{[t]}) \le \lambda_{\max}((\hat{\boldsymbol{\Sigma}}_{\mathcal{A}})_{\mathcal{S}',\mathcal{S}'}^{[t]}) \le \overline{\gamma}, \forall \mathcal{S}' \subseteq [p] \text{ with } |\mathcal{S}'| \le 3s', \forall t = 0,\dots,T-1\} \\ & \bigcap \{\|(\hat{\boldsymbol{\Sigma}}_{\mathcal{A}})_{\mathcal{S}',\mathcal{S}'}^{[t]} - (\boldsymbol{\Sigma}_{\mathcal{A}})_{\mathcal{S}',\mathcal{S}'}\|_2 \le C\sqrt{\frac{s'\log(p/\eta)}{N/T}}, \forall \mathcal{S}' \subseteq [p] \text{ with } |\mathcal{S}'| \le 3s', \forall t = 0,\dots,T-1\}.\end{aligned}$$

According to Lemma 23 in Li et al. (2024), by choosing $\underline{\gamma} = 10L_0^{-1}/11$ and $\overline{\gamma} = 10L_0/9$, then for sufficient large $C > 0$, it holds that $\mathbb{P}(\mathcal{E}_1) \ge 1 - \eta/4$.

Denote $\boldsymbol{\Delta}^t = \sum_{k\in\{0\}\cup\mathcal{A}} (n_k/N)\boldsymbol{\Delta}_t^{(k)}$ as the total noise added in gradient aggregation step at the t-th iteration. We define

$$\begin{aligned}\mathcal{E}_2 = & \left\{ \|\boldsymbol{\Delta}^t\|_\infty^2 \le C\frac{T^2 R_d^2 \log(1/\delta)\log(Tp/\eta)}{N^2\epsilon^2} \sum_{k\in\{0\}\cup\mathcal{A}} R_k^2, \quad \forall t \in \{0,\dots,T-1\} \right\} \\ & \bigcap \left\{ \|\boldsymbol{\Delta}_Y\|_\infty \le C\frac{R_Y\sqrt{\log(n_0/\eta)\log(1/\delta)}}{\epsilon} \right\}.\end{aligned}$$

Notice that due to the sub-Gaussian tail bound, for sufficient large $C > 0$, it holds that $\mathbb{P}(\mathcal{E}_2) \ge 1 - \eta/4$, which ensures the upper bound of the noise term added in. Indeed, since

$$\boldsymbol{\Delta}^t = \sum_{k\in\{0\}\cup\mathcal{A}} \frac{n_k}{N}\boldsymbol{\Delta}_t^{(k)}, \qquad \boldsymbol{\Delta}_t^{(k)} \sim \mathrm{N}\left(\mathbf{0}, \frac{8T^2R_k^2R_d^2\log(c/\delta)}{n_k^2(\epsilon/2)^2}\mathbf{I}_p\right),$$

we have, for each coordinate $j \in [p]$,

$$\mathrm{Var}(\Delta_j^t) = \sum_{k\in\{0\}\cup\mathcal{A}} \frac{n_k^2}{N^2}\frac{8T^2R_k^2R_d^2\log(c/\delta)}{n_k^2(\epsilon/2)^2} \lesssim \frac{T^2R_d^2\log(1/\delta)}{N^2\epsilon^2}\sum_{k\in\{0\}\cup\mathcal{A}} R_k^2.$$

Therefore, by a Gaussian maximal inequality and a union bound over $j \in [p]$ and $t \in \{0,\dots,T-1\}$, the first part of $\mathcal{E}_2$ holds with probability at least $1-\eta/4$.

Then we define the event

$$\mathcal{E}_3 = \left\{\frac{T}{N}\left\|\sum_{k\in\mathcal{A}}(\boldsymbol{X}^{(k)t})^\top\boldsymbol{\varepsilon}^{(k)t}\right\|_\infty \le C\sqrt{\frac{\log(Tp/\eta)}{N/T}}, \quad \forall t=0,\dots,T-1\right\}$$
$$\bigcap\left\{\|\boldsymbol{Y}^{(0)} - \hat{\boldsymbol{m}}^{(init)}(\boldsymbol{W}^{(0)})\|_n \le 50\sqrt{\frac{\log(p/\eta)}{n_0}}\right\}.$$

According to Bernstein's inequality and Lemma S.2 in Zhu (2017), we have sufficient large $C>0$, such that $\mathbb{P}(\mathcal{E}_3) \ge 1-\eta/4$. The final event is introduced to ensure that truncation occurs only rarely across all iterations. Notice that the last sub-event in event $\mathcal{E}_3$ implies $\{\|\boldsymbol{Y}^{(0)} - \hat{\boldsymbol{m}}^{(init)}(\boldsymbol{W}^{(0)})\|_2 \le 50\sqrt{\log(p/\eta)}\}$, which indicates a rare truncation happens within the residulization step. Define the event $\mathcal{E}_4$ as

$$\begin{aligned}\mathcal{E}_4 = &\{R_0 \ge |(\boldsymbol{X}_i^{(0)})^\top\boldsymbol{\beta}^t - \breve{Y}_i^{(0)}|, \forall i\in\mathcal{I}^{(0)[t]}, \forall t=0,\dots,T-1\}\\ &\bigcap\{R_k \ge |(\boldsymbol{X}_i^{(k)})^\top\boldsymbol{\beta}^t - Y_i^{(k)}|, \forall i\in\mathcal{I}^{(k)[t]}, \forall k\in\mathcal{A}, \forall t=0,\dots,T-1\}\\ &\bigcap\{R_d \ge \|\boldsymbol{X}_i^{(k)}\|_2, \forall i\in\mathcal{I}^{(k)[t]}, \forall k\in\{0\}\cup\mathcal{A}, \forall t=0,\dots,T-1\},\end{aligned}$$

where $R_0 = s'R_d + \sqrt{\log(n_0/\eta)\log(1/\delta)}R_Y/\epsilon$, $R_k = 2(\log(18n_k/\eta)(\sigma^2+h))^{1/2}$ and $R_d = (p\log(18N/\eta))^{1/2}\tilde{\sigma}$, as defined in Theorem 1. According to the union bound and the sub-Gaussian tail bounds, it can be verified that $\mathbb{P}(\mathcal{E}_4) \ge 1-\eta/4$.

Combining the results in the above four events, we conclude the first result:

$$\mathbb{P}(\cap_{i=1}^4\mathcal{E}_i) \ge 1-\eta. \tag{1}$$

For the gradient release, fix $k \in \{0\}\cup\mathcal{A}$ and $t\in\{0,\dots,T-1\}$. Let $n_{k,t} = |\mathcal{I}^{(k)[t]}|$ and define the clipped block-gradient query

$$\boldsymbol{G}_{k,t}(D_k) = \frac{1}{n_{k,t}}\sum_{i\in\mathcal{I}^{(k)[t]}}\Pi_{R_k}\{(\boldsymbol{X}_i^{(k)})^\top\boldsymbol{\beta}^t - \breve{Y}_i^{(k)}\}\Pi_{R_d}(\boldsymbol{X}_i^{(k)}).$$

By construction,

$$\left|\Pi_{R_k}\{(\boldsymbol{X}_i^{(k)})^\top\boldsymbol{\beta}^t - \breve{Y}_i^{(k)}\}\right| \le R_k, \qquad \left\|\Pi_{R_d}(\boldsymbol{X}_i^{(k)})\right\|_2 \le R_d.$$

Thus each summand has Euclidean norm at most $R_k R_d$. For two neighboring local datasets D_k and D_k', only one summand in the corresponding block can change. Hence

$$\|\boldsymbol{G}_{k,t}(D_k) - \boldsymbol{G}_{k,t}(D_k')\|_2 \leq \frac{1}{n_{k,t}}\{R_k R_d + R_k R_d\} = \frac{2R_k R_d}{n_{k,t}} \leq \frac{2TR_k R_d}{n_k}.$$

Therefore, the ℓ_2-sensitivity of the clipped block-gradient query is bounded by

$$\Delta_{k,t} \leq \frac{2TR_k R_d}{n_k}.$$

Consequently, by the Gaussian mechanism, adding noise with covariance

$$\frac{2\log(5/\delta)}{(\epsilon/2)^2}\left(\frac{2TR_k R_d}{n_k}\right)^2 \mathbf{I}_p = \frac{8T^2 R_k^2 R_d^2 \log(5/\delta)}{n_k^2(\epsilon/2)^2}\mathbf{I}_p$$

ensures $(\epsilon/2, \delta/2)$-DP for the t-th block-gradient release, conditional on the previous private transcript. The deterministic factor n_k/N in the aggregated gradient is post-processing and does not affect the privacy guarantee.

Since the blocks $\{\mathcal{I}^{(k)[t]}\}_{t=0}^{T-1}$ are mutually disjoint for each site k, parallel composition implies that the full gradient-release transcript over $t = 0, \ldots, T-1$ remains $(\epsilon/2, \delta/2)$-DP at each site. It remains to verify the target residualization release. Although $\hat{\boldsymbol{m}}^{(init)}(\boldsymbol{W}^{(0)})$ is fitted with the full target sample, the released residualization vector is projected onto the ℓ_2-ball with radius R_Y. Hence its ℓ_2-sensitivity satisfies

$$\begin{aligned}
&\sup_{\mathcal{D}_0, \mathcal{D}_0'} \left\| \Pi_{R_Y}\{\boldsymbol{Y}^{(0)} - \hat{\boldsymbol{m}}^{(init)}(\boldsymbol{W}^{(0)})\} - \Pi_{R_Y}\{(\boldsymbol{Y}^{(0)})' - \hat{\boldsymbol{m}}^{(init)\prime}((\boldsymbol{W}^{(0)})')\} \right\|_2 \\
&\leq \sup_{\mathcal{D}_0, \mathcal{D}_0'} \left[\left\| \Pi_{R_Y}\{\boldsymbol{Y}^{(0)} - \hat{\boldsymbol{m}}^{(init)}(\boldsymbol{W}^{(0)})\} \right\|_2 + \left\| \Pi_{R_Y}\{(\boldsymbol{Y}^{(0)})' - \hat{\boldsymbol{m}}^{(init)\prime}((\boldsymbol{W}^{(0)})')\} \right\|_2 \right] \leq 2R_Y,
\end{aligned}$$

where the supremum is taken over neighboring target datasets $\mathcal{D}_0$ and $\mathcal{D}_0'$. Thus the ℓ_2-sensitivity of the residualization query is at most $2R_Y$, and the Gaussian perturbation

$$\boldsymbol{\Delta}_Y \sim \mathrm{N}\left(\mathbf{0}, \frac{2\log(5/\delta)(2R_Y)^2}{(\epsilon/2)^2}\mathbf{I}_{n_0}\right)$$

ensures $(\epsilon/2, \delta/2)$-DP for this target residualization step. Therefore, by sequential composition between the target residualization release and the gradient-release transcript, Algorithm 1 satisfies (ϵ, δ)-STDP. Notice that $\eta > 0$ can be arbitrary small in (1), hence the remaining part of the proof will be processed conditional on the event $\cap_{i=1}^4 \mathcal{E}_i$.

(II) Now we turn to the second part of the proof to establish consistency. Let the loss function of linear part be

$$\mathcal{L}_N^t(\boldsymbol{\beta}) = \frac{1}{2}\sum_{k\in\{0\}\cup\mathcal{A}} \frac{n_k}{N}\frac{1}{n_{k,t}} \sum_{i\in\mathcal{I}^{(k)[t]}} [\breve{Y}_i^{(k)} - (\boldsymbol{X}_i^{(k)})^\top \boldsymbol{\beta}]^2.$$

When the blocks have equal sizes, $n_{k,t} = n_k/T$, this is equivalently

$$\mathcal{L}_N^t(\boldsymbol{\beta}) = \frac{1}{2(N/T)} \sum_{k\in\{0\}\cup\mathcal{A}} \sum_{i\in\mathcal{I}^{(k)[t]}} [\breve{Y}_i^{(k)} - (\boldsymbol{X}_i^{(k)})^\top \boldsymbol{\beta}]^2.$$

The gradient is

$$\boldsymbol{G}^t = \nabla \mathcal{L}_N^t(\boldsymbol{\beta}^t).$$

In order to show the consistency, firstly, we will show the upper bound of $\mathcal{L}_N^t(\boldsymbol{\beta}^{t+1}) - \mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})$ consists of $\mathcal{L}_N^t(\boldsymbol{\beta}^t) - \mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})$, the difference $\boldsymbol{\beta}^t - \boldsymbol{\beta}^{(0)}$, and the infinity norm $\|\boldsymbol{\Delta}^t\|_\infty$. Second, we derive a lower bound for $\mathcal{L}_N^t(\boldsymbol{\beta}^{t+1}) - \mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})$ in terms of $\|\boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^{(0)}\|_2$, and an upper bound for $\mathcal{L}_N^t(\boldsymbol{\beta}^t) - \mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})$ in terms of $\|\boldsymbol{\beta}^t - \boldsymbol{\beta}^{(0)}\|_2$. Finally, we simplify these bounds to establish an inductive relationship between $\|\boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}}$ and $\|\boldsymbol{\beta}^t - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}}$, translate the $\boldsymbol{\Sigma}$-norm to the ℓ_2-norm and complete the proof. Throughout this part, we will use C as some universal positive constant and its value may vary from line to line. We also use C with different suffix such as C_1 and C_2 to represent different constant within a line, but they still may also vary across different lines. All the constants are irrelative to the sample size n_k, the dimension p and the sparsity level s, s'.

Firstly, we define

$$\tilde{\mathcal{S}}^t = \mathcal{S}^{t+1} \cup \mathcal{S}^t \cup \mathcal{S}, \tag{2}$$

where $\mathcal{S}$, $\mathcal{S}^t$ and $\mathcal{S}^{t+1}$ are the non-zero support for $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\beta}^t$ and $\boldsymbol{\beta}^{t+1}$, respectively. We start with applying Taylor's expansion as it holds that

$$\begin{aligned}
&\mathcal{L}_N^t(\boldsymbol{\beta}^{t+1}) - \mathcal{L}_N^t(\boldsymbol{\beta}^t) &(3)\\
=&\langle \boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^t, \boldsymbol{G}^t\rangle + \frac{1}{2}(\boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^t)^\top \nabla^2 \mathcal{L}_N^t(\tilde{\boldsymbol{\beta}})(\boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^t)\\
\leq&\langle \boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^t, \boldsymbol{G}^t\rangle + \frac{1}{2}\overline{\gamma}\|\boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^t\|_2^2\\
=&\frac{1}{2}\overline{\gamma}\|\boldsymbol{\beta}^{t+1}_{\tilde{\mathcal{S}}^t} - \boldsymbol{\beta}^t_{\tilde{\mathcal{S}}^t} + \xi/\overline{\gamma}\cdot \boldsymbol{G}^t_{\tilde{\mathcal{S}}^t}\|_2^2 - \frac{\xi^2}{2\overline{\gamma}}\|\boldsymbol{G}^t_{\tilde{\mathcal{S}}^t}\|_2^2 + (1-\xi)\langle \boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^t, \boldsymbol{G}^t\rangle, &(4)
\end{aligned}$$

where $\tilde{\boldsymbol{\beta}}$ is the intermediate term between $\boldsymbol{\beta}^{t+1}$ and $\boldsymbol{\beta}^t$. The inequality is due to Lemma 23.(ii) in Li et al. (2024) and the last equality is generally true for any real number $\xi \in \mathbb{R}$. For technical reason, we specify $\xi = 1 - 0.296/L_0^4$ in this proof. The last term in (3) can be bounded as

$$
\begin{aligned}
&\langle \boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^t, \boldsymbol{G}^t \rangle \\
=&\langle \boldsymbol{\beta}^{t+1}_{\mathcal{S}^{t+1}} - \boldsymbol{\beta}^t_{\mathcal{S}^{t+1}}, \boldsymbol{G}^t_{\mathcal{S}^{t+1}} \rangle - \langle \boldsymbol{\beta}^t_{\mathcal{S}^t \setminus \mathcal{S}^{t+1}}, \boldsymbol{G}^t_{\mathcal{S}^t \setminus \mathcal{S}^{t+1}} \rangle \\
=&\langle -\rho_1 \boldsymbol{G}^t_{\mathcal{S}^{t+1}} - \rho_1 \boldsymbol{\Delta}^t_{\mathcal{S}^{t+1}}, \boldsymbol{G}^t_{\mathcal{S}^{t+1}} \rangle - \langle \boldsymbol{\beta}^t_{\mathcal{S}^t \setminus \mathcal{S}^{t+1}}, \boldsymbol{G}^t_{\mathcal{S}^t \setminus \mathcal{S}^{t+1}} \rangle \\
\overset{(a)}{\le}& -\rho_1 \|\boldsymbol{G}^t_{\mathcal{S}^{t+1}}\|_2^2 + 10\rho_1 \|\boldsymbol{\Delta}^t_{\mathcal{S}^{t+1}}\|_2^2 + \frac{\rho_1}{40} \|\boldsymbol{G}^t_{\mathcal{S}^{t+1}}\|_2^2 - \langle \boldsymbol{\beta}^t_{\mathcal{S}^t \setminus \mathcal{S}^{t+1}}, \boldsymbol{G}_{\mathcal{S}^t \setminus \mathcal{S}^{t+1}} \rangle \\
\le& -\rho_1 \|\boldsymbol{G}^t_{\mathcal{S}^{t+1}}\|_2^2 + 10 s_0 \rho_1 \|\boldsymbol{\Delta}^t\|_\infty^2 + \frac{\rho_1}{40} \|\boldsymbol{G}^t_{\mathcal{S}^{t+1}}\|_2^2 - \langle \boldsymbol{\beta}^t_{\mathcal{S}^t \setminus \mathcal{S}^{t+1}}, \boldsymbol{G}_{\mathcal{S}^t \setminus \mathcal{S}^{t+1}} \rangle, \qquad (5)
\end{aligned}
$$

where (a) uses the Cauchy-Schwarz's inequality. Apply Lemma 2 for $\langle \boldsymbol{\beta}^t_{\mathcal{S}^t \setminus \mathcal{S}^{t+1}}, \boldsymbol{G}_{\mathcal{S}^t \setminus \mathcal{S}^{t+1}} \rangle$ and combine (5), we obtain

$$\langle \boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^t, \boldsymbol{G}^t \rangle \le -\frac{9}{20} \rho_1 \|\boldsymbol{G}^t_{\mathcal{S}^{t+1} \cup \mathcal{S}^t}\|_2^2 + Cs' \|\boldsymbol{\Delta}^t\|_\infty^2.$$

The second term in (3) can be handled as follows. Consider a set $\mathcal{S}' \subseteq \mathcal{S}^t \setminus \mathcal{S}^{t+1}$ with $|\mathcal{S}'| = |\tilde{\mathcal{S}}^t \setminus (\mathcal{S}^t \cup \mathcal{S})| = |\mathcal{S}^{t+1} \setminus (\mathcal{S}^t \cup \mathcal{S})|$. After applying Lemma 2 again, we can obtain

$$\|\boldsymbol{\beta}^t_{\mathcal{S}'} - \xi/\overline{\gamma} \cdot \boldsymbol{G}^t_{\mathcal{S}'}\|_2^2 \le \frac{1+c}{1-c} \frac{\xi^2}{\overline{\gamma}^2} \|\boldsymbol{G}^t_{\mathcal{S}^{t+1} \setminus (\mathcal{S}^t \cup \mathcal{S})}\|_2^2 + \frac{1}{1-c} (1 + 2/c) s' \|\boldsymbol{\Delta}^t\|_\infty^2, \qquad (6)$$

which entails that

$$-\frac{\xi^2}{\overline{\gamma}} \|\boldsymbol{G}^t_{\mathcal{S}^{t+1} \setminus (\mathcal{S}^t \cup \mathcal{S})}\|_2^2 \le -\frac{1-c}{1+c} \overline{\gamma} \|\boldsymbol{\beta}^t_{\mathcal{S}'} - \xi/\overline{\gamma} \cdot \boldsymbol{G}^t_{\mathcal{S}'}\|_2^2 + \overline{\gamma} \frac{1 + 2/c}{1+c} s' \|\boldsymbol{\Delta}^t\|_\infty^2.$$

The first two terms in (5) can thus be bounded as

$$\begin{aligned}
&\frac{1}{2}\overline{\gamma}\Big\|\boldsymbol{\beta}^{t+1}_{\tilde{\mathcal{S}}^t}-\boldsymbol{\beta}^{t}_{\tilde{\mathcal{S}}^t}+\tfrac{\xi}{\overline{\gamma}}\,\boldsymbol{G}^{t}_{\tilde{\mathcal{S}}^t}\Big\|_2^2-\frac{\xi^2}{2\overline{\gamma}}\Big\|\boldsymbol{G}^{t}_{\tilde{\mathcal{S}}^t\setminus(\mathcal{S}^t\cup\mathcal{S})}\Big\|_2^2\\
=&\frac{1}{2}\,\overline{\gamma}\Big\|\big(\mathrm{HT}(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t,s')\big)_{\tilde{\mathcal{S}}^t}-\boldsymbol{\beta}^{t}_{\tilde{\mathcal{S}}^t}+\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^{t}_{\tilde{\mathcal{S}}^t}\Big\|_2^2-\frac{\xi^2}{2\overline{\gamma}}\Big\|\boldsymbol{G}^{t}_{\tilde{\mathcal{S}}^t\setminus(\mathcal{S}^t\cup\mathcal{S})}\Big\|_2^2\\
\overset{(a)}{\leq}&C\,\overline{\gamma}\,\big(\tfrac{\xi}{\overline{\gamma}}\big)^2\|\boldsymbol{\Delta}^t_{\tilde{\mathcal{S}}^t}\|_2^2+\frac{\overline{\gamma}}{2}\,\frac{1+c}{1-c}\,\Big\|\big(\mathrm{HT}(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t,s')\big)_{\tilde{\mathcal{S}}^t}-(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\tilde{\mathcal{S}}^t}\Big\|_2^2\\
&\quad-\frac{\overline{\gamma}}{2}\,\frac{1-c}{1+c}\,\|\boldsymbol{\beta}^t_{\mathcal{S}'}-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t_{\mathcal{S}'}\|_2^2+\frac{\overline{\gamma}}{2}\,\frac{1+2/c}{1+c}\,s'\,\big(\tfrac{\xi}{\overline{\gamma}}\big)^2\|\boldsymbol{\Delta}^t\|_\infty^2\\
\overset{(b)}{\leq}&C\,\overline{\gamma}\,s'\,\big(\tfrac{\xi}{\overline{\gamma}}\big)^2\|\boldsymbol{\Delta}^t\|_\infty^2+\frac{\overline{\gamma}}{2}\,\frac{1+c}{1-c}\,\Big\|\big(\mathrm{HT}(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t,s')\big)_{\tilde{\mathcal{S}}^t}-(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\tilde{\mathcal{S}}^t}\Big\|_2^2\\
&\quad-\frac{\overline{\gamma}}{2}\Big(\frac{1-c}{1+c}\Big)^2\Big\|[\mathrm{HT}(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t,s')]_{\mathcal{S}'}-(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\mathcal{S}'}\Big\|_2^2\\
\overset{(c)}{\leq}&C\,s'\,\big(\tfrac{\xi}{\overline{\gamma}}\big)^2\|\boldsymbol{\Delta}^t\|_\infty^2+\frac{5\overline{\gamma}}{9}\Big\|\big(\mathrm{HT}(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t,s')\big)_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}-(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}\Big\|_2^2\\
&\quad+\frac{(1+c)^3-(1-c)^3}{(1+c)^2(1-c)}\,\frac{\overline{\gamma}}{2}\|(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\mathcal{S}'}\|_2^2 \qquad (7)\\
\overset{(d)}{\leq}&C\,s'\,\big(\tfrac{\xi}{\overline{\gamma}}\big)^2\|\boldsymbol{\Delta}^t\|_\infty^2+\frac{5\overline{\gamma}}{9}\Big\|\big(\mathrm{HT}(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t,s')\big)_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}-(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}\Big\|_2^2\\
&\quad+\frac{c^3+3c}{(1+c)(1-c)^2}\,\frac{\xi^2}{\overline{\gamma}}\,\|\boldsymbol{G}^t_{\mathcal{S}^{t+1}\setminus(\mathcal{S}^t\cup\mathcal{S})}\|_2^2 \qquad (8)\\
\leq&C\,\overline{\gamma}\,s'\,\big(\tfrac{\xi}{\overline{\gamma}}\big)^2\|\boldsymbol{\Delta}^t\|_\infty^2+\frac{5\overline{\gamma}}{9}\,\frac{s_0}{s'}\Big\|\boldsymbol{\beta}^{(0)}_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}-\boldsymbol{\beta}^t_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}+\tfrac{\xi}{\overline{\gamma}}\,\boldsymbol{G}^t_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}\Big\|_2^2+\frac{c^3+3c}{(1+c)(1-c)^2}\,\frac{\xi^2}{\overline{\gamma}}\|\boldsymbol{G}^t_{\mathcal{S}^{t+1}\setminus(\mathcal{S}^t\cup\mathcal{S})}\|_2^2\\
\leq&C\,\overline{\gamma}\,s'\,\big(\tfrac{\xi}{\overline{\gamma}}\big)^2\|\boldsymbol{\Delta}^t\|_\infty^2+\frac{5\overline{\gamma}}{9}\,\frac{s_0}{s'}\Big\|\boldsymbol{\beta}^{(0)}_{\tilde{\mathcal{S}}^t}-\boldsymbol{\beta}^t_{\tilde{\mathcal{S}}^t}+\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t_{\tilde{\mathcal{S}}^t}\Big\|_2^2+\frac{c^3+3c}{(1+c)(1-c)^2}\,\frac{\xi^2}{\overline{\gamma}}\|\boldsymbol{G}^t_{\mathcal{S}^{t+1}\setminus(\mathcal{S}^t\cup\mathcal{S})}\|_2^2\\
\leq&C\,\overline{\gamma}\,s'\,\big(\tfrac{\xi}{\overline{\gamma}}\big)^2\|\boldsymbol{\Delta}^t\|_\infty^2+\frac{5}{9}\,\frac{s_0}{s'}\Big(2\xi\langle\boldsymbol{\beta}^{(0)}_{\tilde{\mathcal{S}}^t}-\boldsymbol{\beta}^t_{\tilde{\mathcal{S}}^t},\boldsymbol{G}^t_{\tilde{\mathcal{S}}^t}\rangle+\overline{\gamma}\|\boldsymbol{\beta}^{(0)}_{\tilde{\mathcal{S}}^t}-\boldsymbol{\beta}^t_{\tilde{\mathcal{S}}^t}\|_2^2+\tfrac{\xi^2}{\overline{\gamma}}\|\boldsymbol{G}^t_{\tilde{\mathcal{S}}^t}\|_2^2\Big)\\
&\quad+\frac{c^3+3c}{(1+c)(1-c)^2}\,\frac{\xi^2}{\overline{\gamma}}\,\|\boldsymbol{G}^t_{\mathcal{S}^{t+1}}\|_2^2\\
\leq&C\,s'\,\big(\tfrac{\xi}{\overline{\gamma}}\big)^2\|\boldsymbol{\Delta}^t\|_\infty^2+\frac{5}{9}\,\frac{s_0}{s'}\Big(2\xi\,\mathcal{L}^t_N(\boldsymbol{\beta}^{(0)})-2\xi\,\mathcal{L}^t_N(\boldsymbol{\beta}^t)+(\overline{\gamma}-\xi\underline{\gamma})\|\boldsymbol{\beta}^{(0)}-\boldsymbol{\beta}^t\|_2^2+\tfrac{\xi^2}{\overline{\gamma}}\|\boldsymbol{G}^t_{\tilde{\mathcal{S}}^t}\|_2^2\Big)\\
&\quad+\frac{c^3+3c}{(1+c)(1-c)^2}\,\frac{\xi^2}{\overline{\gamma}}\,\|\boldsymbol{G}^t_{\mathcal{S}^{t+1}}\|_2^2. \qquad (9)
\end{aligned}$$

Here, (a) and (b) are based on Cauchy-Schwarz's inequality and Lemma 2. Next, (c) holds according to the fact that $\Big\|\big(\mathrm{HT}(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t,s')\big)_{\tilde{\mathcal{S}}^t}-(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\tilde{\mathcal{S}}^t}\Big\|_2^2-\Big\|\big(\mathrm{HT}(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t,s')\big)_{\mathcal{S}'}-(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\mathcal{S}'}\Big\|_2^2=\Big\|\big(\mathrm{HT}(\boldsymbol{\beta}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{G}^t-\tfrac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t,s')\big)_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}-$

$(\boldsymbol{\beta}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{G}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}\Big\|_2^2$. Together with (6), it leads to (d). Applying the Lemma 21 in Li et al. (2024) to $(\mathrm{HT}(\boldsymbol{\beta}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{G}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t, s'))_{\tilde{\mathcal{S}}^t} - (\boldsymbol{\beta}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{G}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\tilde{\mathcal{S}}^t}$, it yields that

$$\begin{aligned}
&\left\|(\mathrm{HT}(\boldsymbol{\beta}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{G}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t, s'))_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'} - (\boldsymbol{\beta}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{G}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}\right\|_2^2 \\
\leq & \frac{|\tilde{\mathcal{S}}^t\setminus\mathcal{S}'| - s'}{|\tilde{\mathcal{S}}^t\setminus\mathcal{S}'| - s_0}\|\boldsymbol{\beta}^{(0)}_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'} - (\boldsymbol{\beta}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{G}^t - \frac{\xi}{\overline{\gamma}}\boldsymbol{\Delta}^t)_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}\|_2^2 \\
\leq & \frac{s_0}{s'}\|\boldsymbol{\beta}^{(0)}_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'} - \boldsymbol{\beta}^t_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'} + \frac{\xi}{\overline{\gamma}}\boldsymbol{G}^t_{\tilde{\mathcal{S}}^t\setminus\mathcal{S}'}\|_2^2 + \frac{s_0}{s'}s'\left(\frac{\xi}{\overline{\gamma}}\right)^2\|\boldsymbol{\Delta}^t\|_\infty^2\,, \qquad (10)
\end{aligned}$$

where the second inequality used the triangle inequality and the fact that $|\tilde{\mathcal{S}}^t\setminus\mathcal{S}'| \leq s_0 + s'$. According to (10) and the convexity of $\mathcal{L}_N^t(\cdot)$, the remaining inequalities hold and we obtain (9).

Now combine (5) and (9), and insert them into (3). Since $\rho_1 = \xi/\overline{\gamma}$, then it holds that

$$\begin{aligned}
&\mathcal{L}_N^t(\boldsymbol{\beta}^{t+1}) - \mathcal{L}_N^t(\boldsymbol{\beta}^t) \\
\leq & \frac{1}{2}\overline{\gamma}\left\|\boldsymbol{\beta}^{t+1}_{\tilde{\mathcal{S}}^t} - \boldsymbol{\beta}^t_{\tilde{\mathcal{S}}^t} + \frac{\xi}{\overline{\gamma}}\cdot\boldsymbol{G}^t_{\tilde{\mathcal{S}}^t}\right\|_2^2 - \frac{\xi^2}{2\overline{\gamma}}\left\|\boldsymbol{G}^t_{\tilde{\mathcal{S}}^t\setminus(\mathcal{S}^t\cup\mathcal{S})}\right\|_2^2 - \frac{\xi^2}{2\overline{\gamma}}\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2^2 \\
&- \frac{9\xi}{20\overline{\gamma}}(1-\xi)\left\|\boldsymbol{G}^t_{\mathcal{S}^{t+1}\cup\mathcal{S}^t}\right\|_2^2 + Cs'\|\boldsymbol{\Delta}^t\|_\infty^2 \\
\leq & \frac{10s_0}{9s'}\cdot\xi\left[\mathcal{L}_N^t(\boldsymbol{\beta}^{(0)}) - \mathcal{L}_N^t(\boldsymbol{\beta}^t)\right] + \frac{s_0}{s'}\cdot\frac{5(\overline{\gamma} - \xi\underline{\gamma})}{9}\left\|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^t\right\|_2^2 \\
&+ \frac{s_0}{s'}\cdot\frac{5\xi^2}{9\overline{\gamma}}\left(\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2^2 + \|\boldsymbol{G}^t_{\mathcal{S}^{t+1}\setminus(\mathcal{S}^t\cup\mathcal{S})}\|_2^2\right) \\
&- \frac{\xi^2}{2\overline{\gamma}}\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2^2 - \frac{9\xi}{20\overline{\gamma}}(1-\xi)\|\boldsymbol{G}^t_{\mathcal{S}^{t+1}\cup\mathcal{S}^t}\|_2^2 + Cs'\|\boldsymbol{\Delta}^t\|_\infty^2 \\
= & \frac{10s_0}{9s'}\cdot\xi\left[\mathcal{L}_N^t(\boldsymbol{\beta}^{(0)}) - \mathcal{L}_N^t(\boldsymbol{\beta}^t)\right] + \frac{s_0}{s'}\cdot\frac{5(\overline{\gamma} - \xi\underline{\gamma})}{9}\|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^t\|_2^2 \\
&+ \left[\frac{s_0}{s'}\cdot\frac{5\xi^2}{9\overline{\gamma}} - \frac{9\xi}{20\overline{\gamma}}(1-\xi)\right]\|\boldsymbol{G}^t_{\mathcal{S}^{t+1}\setminus(\mathcal{S}^t\cup\mathcal{S})}\|_2^2 \\
&+ \left(\frac{10s_0}{9s'} - 1\right)\cdot\frac{\xi^2}{2\overline{\gamma}}\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2^2 + Cs'\|\boldsymbol{\Delta}^t\|_\infty^2 \\
\leq & \frac{10s_0}{9s'}\cdot\xi\left[\mathcal{L}_N^t(\boldsymbol{\beta}^{(0)}) - \mathcal{L}_N^t(\boldsymbol{\beta}^t)\right] + \frac{s_0}{s'}\cdot\frac{5(\overline{\gamma} - \xi\underline{\gamma})}{9}\|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^t\|_2^2 \\
&- \frac{9s' - 10s_0}{9s'}\cdot\frac{\xi^2}{2\overline{\gamma}}\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2^2 + Cs'\|\boldsymbol{\Delta}^t\|_\infty^2. \qquad (11)
\end{aligned}$$

The last inequality holds because of $\frac{s_0}{s'} \cdot \frac{5\xi^2}{9\overline{\gamma}} - \frac{9\xi}{20\overline{\gamma}}(1-\xi) \leq 0$, which is due to the condition $4.18 L_0^4 s_0 \leq s'$ in Theorem 1. The remaining term we have to bound in (11) is $\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2^2$. According to the convexity of $\mathcal{L}_N^t(\cdot)$, we have

$$\begin{aligned}\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2^2 - \frac{1}{4}\underline{\gamma}^2\|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^t\|_2^2 &= (\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2 + \frac{\underline{\gamma}}{2}\|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^t\|_2)(\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2 - \frac{\underline{\gamma}}{2}\|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^t\|_2) \\ &\geq \frac{\mathcal{L}_N^t(\boldsymbol{\beta}^t) - \mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})}{\|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^t\|_2} \cdot (\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2 + \frac{\underline{\gamma}}{2}\|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^t\|_2),\end{aligned}$$

which implies

$$\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2^2 \geq \frac{1}{4}\underline{\gamma}^2\|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^t\|_2^2 + \frac{\underline{\gamma}}{2}[\mathcal{L}_N^t(\boldsymbol{\beta}^t) - \mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})]. \tag{12}$$

By adding $\mathcal{L}_N^t(\boldsymbol{\beta}^t) - \mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})$ on both sides of (11), together with (12), we obtain

$$\begin{aligned}\mathcal{L}_N^t(\boldsymbol{\beta}^{t+1}) - \mathcal{L}_N^t(\boldsymbol{\beta}^{(0)}) \leq& (1 - \frac{10s_0}{9s'}\cdot\xi)\cdot[\mathcal{L}_N^t(\boldsymbol{\beta}^t) - \mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})] + \frac{s_0}{s'}\cdot\frac{5(\overline{\gamma} - \xi\underline{\gamma})}{9}\|\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}^t\|_2^2 \\ & - \frac{9s' - 10s_0}{9s'}\cdot\frac{\xi^2}{2\overline{\gamma}}\cdot\|\boldsymbol{G}^t_{\mathcal{S}^t\cup\mathcal{S}}\|_2^2 + Cs'\|\boldsymbol{\Delta}^t\|_\infty^2 \\ \leq& (1 - \frac{10s_0}{9s'}\xi - \frac{9s' - 10s_0}{9s'}\cdot\frac{\xi^2}{4\overline{\gamma}}\underline{\gamma})\cdot[\mathcal{L}_N^t(\boldsymbol{\beta}^t) - \mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})] \\ & + (\frac{s_0}{s'}\cdot\frac{5(\overline{\gamma} - \xi\underline{\gamma})}{9} - \frac{9s' - 10s_0}{9s'}\cdot\frac{\xi^2}{8\overline{\gamma}}\underline{\gamma}^2)\|\boldsymbol{\beta}^t - \boldsymbol{\beta}^{(0)}\|_2^2 + Cs'\|\boldsymbol{\Delta}^t\|_\infty^2.\end{aligned} \tag{13}$$

Now we need to derive a lower bound involving $\|\boldsymbol{\beta}^{t+1}-\boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}$ for $\mathcal{L}_N^t(\boldsymbol{\beta}^{t+1})-\mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})$ and an upper bound involving $\|\boldsymbol{\beta}^{t}-\boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}$ for $\mathcal{L}_N^t(\boldsymbol{\beta}^{t})-\mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})$. For any matrix $\boldsymbol{\Sigma}$, define

$$\|\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}}^2=(\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)})^\top\boldsymbol{\Sigma}(\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)}).$$

Let

$$\mathcal{S}_+^t=\mathcal{S}^t\cup\mathcal{S},\qquad \boldsymbol{r}^{(0)t}=\boldsymbol{Y}^{(0)t}-\boldsymbol{X}^{(0)t}\boldsymbol{\beta}^{(0)}-\hat{\boldsymbol{m}}^{(init)}(\boldsymbol{W}^{(0)t}),$$

and let $\boldsymbol{\Delta}_Y^t$ denote the sub-vector of $\boldsymbol{\Delta}_Y$ corresponding to the target block $\mathcal{I}^{(0)[t]}$. Then, on the high-probability event where the projection in the residualization step is inactive,

$$\breve{\boldsymbol{Y}}^{(0)t}=\boldsymbol{X}^{(0)t}\boldsymbol{\beta}^{(0)}+\boldsymbol{r}^{(0)t}+\boldsymbol{\Delta}_Y^t.$$

Therefore,

$$\begin{aligned}
&\mathcal{L}_N^t(\boldsymbol{\beta}^t)-\mathcal{L}_N^t(\boldsymbol{\beta}^{(0)})\\
=&\frac{1}{2(N/T)}\sum_{k\in\{0\}\cup\mathcal{A}}\|\breve{\boldsymbol{Y}}^{(k)t}-\boldsymbol{X}^{(k)t}\boldsymbol{\beta}^t\|_2^2-\frac{1}{2(N/T)}\sum_{k\in\{0\}\cup\mathcal{A}}\|\breve{\boldsymbol{Y}}^{(k)t}-\boldsymbol{X}^{(k)t}\boldsymbol{\beta}^{(0)}\|_2^2\\
=&\frac{1}{2}\|\boldsymbol{\beta}^t-\boldsymbol{\beta}^{(0)}\|_{\widehat{\boldsymbol{\Sigma}}_{\mathcal{A}}^{[t]}}^2+\frac{1}{N/T}(\boldsymbol{\beta}^{(0)}-\boldsymbol{\beta}^t)^\top\sum_{k\in\mathcal{A}}(\boldsymbol{X}_{:,\mathcal{S}_+^t}^{(k)t})^\top\boldsymbol{\varepsilon}^{(k)t}\\
&+\sum_{k\in\mathcal{A}}\frac{n_k}{N}(\boldsymbol{\beta}^{(k)}-\boldsymbol{\beta}^{(0)})^\top\widehat{\boldsymbol{\Sigma}}^{(k)[t]}(\boldsymbol{\beta}^{(0)}-\boldsymbol{\beta}^t)\\
&+\frac{1}{N/T}(\boldsymbol{\beta}^{(0)}-\boldsymbol{\beta}^t)^\top(\boldsymbol{X}_{:,\mathcal{S}_+^t}^{(0)t})^\top\boldsymbol{r}^{(0)t}+\frac{1}{N/T}(\boldsymbol{\beta}^{(0)}-\boldsymbol{\beta}^t)^\top(\boldsymbol{X}_{:,\mathcal{S}_+^t}^{(0)t})^\top\boldsymbol{\Delta}_Y^t\\
\leq&\frac{1}{2}\|\boldsymbol{\beta}^t-\boldsymbol{\beta}^{(0)}\|_{\widehat{\boldsymbol{\Sigma}}_{\mathcal{A}}^{[t]}}^2+\frac{1}{N/T}\|\boldsymbol{\beta}^t-\boldsymbol{\beta}^{(0)}\|_2\left\|\sum_{k\in\mathcal{A}}(\boldsymbol{X}_{:,\mathcal{S}_+^t}^{(k)t})^\top\boldsymbol{\varepsilon}^{(k)t}\right\|_2\\
&+\sum_{k\in\mathcal{A}}\frac{n_k}{N}\|\boldsymbol{\beta}^{(k)}-\boldsymbol{\beta}^{(0)}\|_{\widehat{\boldsymbol{\Sigma}}^{(k)[t]}}\|\boldsymbol{\beta}^t-\boldsymbol{\beta}^{(0)}\|_{\widehat{\boldsymbol{\Sigma}}^{(k)[t]}}\\
&+\frac{1}{N/T}\|\boldsymbol{\beta}^{(0)}-\boldsymbol{\beta}^t\|_2\left\|(\boldsymbol{X}_{:,\mathcal{S}_+^t}^{(0)t})^\top\boldsymbol{r}^{(0)t}\right\|_2\\
&+\frac{1}{N/T}\|\boldsymbol{\beta}^{(0)}-\boldsymbol{\beta}^t\|_2\left\|(\boldsymbol{X}_{:,\mathcal{S}_+^t}^{(0)t})^\top\boldsymbol{\Delta}_Y^t\right\|_2.
\end{aligned}$$

By the standard concentration arguments and the event $\mathcal{E}_1$, we have

$$\left|\frac{1}{2}\|\boldsymbol{\beta}^t-\boldsymbol{\beta}^{(0)}\|^2_{\widehat{\boldsymbol{\Sigma}}^{[t]}_{\mathcal{A}}}-\frac{1}{2}\|\boldsymbol{\beta}^t-\boldsymbol{\beta}^{(0)}\|^2_{\boldsymbol{\Sigma}_{\mathcal{A}}}\right| \lesssim \|\boldsymbol{\beta}^t-\boldsymbol{\beta}^{(0)}\|_2^2\sqrt{\frac{s'\log(p/\eta)}{N/T}},$$
$$\frac{T}{N}\left\|\sum_{k\in\mathcal{A}}(\boldsymbol{X}^{(k)t}_{:,\mathcal{S}^t_+})^\top\boldsymbol{\varepsilon}^{(k)t}\right\|_2 \lesssim \sqrt{\frac{s'\log(p/\eta)}{N/T}},$$
$$\frac{T}{N}\left\|(\boldsymbol{X}^{(0)t}_{:,\mathcal{S}^t_+})^\top\boldsymbol{r}^{(0)t}\right\|_2 \lesssim \sqrt{\frac{s'd\log(p/\eta)}{N/T}},$$
$$\sum_{k\in\mathcal{A}}\frac{n_k}{N}\|\boldsymbol{\beta}^{(k)}-\boldsymbol{\beta}^{(0)}\|_{\widehat{\boldsymbol{\Sigma}}^{(k)[t]}}\|\boldsymbol{\beta}^t-\boldsymbol{\beta}^{(0)}\|_{\widehat{\boldsymbol{\Sigma}}^{(k)[t]}} \lesssim \|\boldsymbol{\beta}^t-\boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}h.$$

Moreover, conditional on the design, since

$$\boldsymbol{\Delta}^t_Y \sim \mathrm{N}(\mathbf{0},\sigma_Y^2\mathbf{I}_{n_0/T}), \qquad \sigma_Y^2=\frac{8\log(5/\delta)R_Y^2}{(\epsilon/2)^2},$$

a Gaussian maximal inequality gives, with probability at least $1-\eta$,

$$\frac{T}{N}\left\|(\boldsymbol{X}^{(0)t}_{:,\mathcal{S}^t_+})^\top\boldsymbol{\Delta}^t_Y\right\|_2 \lesssim \frac{T}{N}\sigma_Y\sqrt{\frac{n_0}{T}s'\log(Tp/\eta)} \lesssim \frac{R_Y\sqrt{n_0Ts'\log(Tp/\eta)\log(5/\delta)}}{N\epsilon}.$$

Therefore, by Cauchy-Schwarz's inequality and the basic fact $ab \le C_4a^2 + C_4^{-1}b^2$, it holds that

$$\mathcal{L}^t_N(\boldsymbol{\beta}^t)-\mathcal{L}^t_N(\boldsymbol{\beta}^{(0)}) \le \left[\frac{1}{2}+C_3\overline{\gamma}\sqrt{\frac{s'\log(p/\eta)}{N/T}}+C_4\right]\|\boldsymbol{\beta}^t-\boldsymbol{\beta}^{(0)}\|^2_{\boldsymbol{\Sigma}_{\mathcal{A}}}+\frac{1}{C_4}h^2 + C_1\frac{s'd\log(p/\eta)}{N/T}+C_7\frac{R_Y^2n_0Ts'\log(Tp/\eta)\log(5/\delta)}{N^2\epsilon^2}. \tag{14}$$

Repeating the same argument for $\mathcal{L}^t_N(\boldsymbol{\beta}^{t+1})-\mathcal{L}^t_N(\boldsymbol{\beta}^{(0)})$, we obtain the corresponding lower bound

$$\mathcal{L}^t_N(\boldsymbol{\beta}^{t+1})-\mathcal{L}^t_N(\boldsymbol{\beta}^{(0)}) \ge \left[\frac{1}{2}-C_5\overline{\gamma}\sqrt{\frac{s'd\log(p/\eta)}{N/T}}-C_6\right]\|\boldsymbol{\beta}^{t+1}-\boldsymbol{\beta}^{(0)}\|^2_{\boldsymbol{\Sigma}_{\mathcal{A}}} - C_1\frac{s'd\log(p/\eta)}{N/T}-\frac{1}{C_6}h^2-C_7\frac{R_Y^2n_0Ts'\log(Tp/\eta)\log(5/\delta)}{N^2\epsilon^2}. \tag{15}$$

Here we choose the constants C_5 and C_6 sufficiently small to enable $\left[\frac{1}{2} - C_5\overline{\gamma}\sqrt{\frac{s'd\log(p/\eta)}{N/T}} - C_6\right] > 0$. Joint with (13), (14) and (15), we obtain an induction relationship between $\|\boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}$ and $\|\boldsymbol{\beta}^{t} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}$ as follows:

$$
\begin{aligned}
&\left[\frac{1}{2} - C_5\overline{\gamma}\sqrt{\frac{s'd\log(p/\eta)}{N/T}} - C_6\right] \|\boldsymbol{\beta}^{t+1} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}^2 \\
\leq & \left(1 - \frac{10s_0}{9s'}\xi - \frac{9s' - 10s_0}{9s'}\frac{\xi^2\underline{\gamma}}{4\overline{\gamma}}\right)\left[\frac{1}{2} + C_3\overline{\gamma}\sqrt{\frac{s'd\log(p/\eta)}{N/T}} + C_4\right] \|\boldsymbol{\beta}^{t} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}^2 \\
&+ \left(\frac{s_0}{s'}\frac{5(\overline{\gamma} - \xi\underline{\gamma})}{9\underline{\gamma}} - \frac{9s' - 10s_0}{9s'}\frac{\xi^2\underline{\gamma}}{8\overline{\gamma}}\right) \|\boldsymbol{\beta}^{t} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}^2 \\
&+ C_0 s'\|\boldsymbol{\Delta}^t\|_\infty^2 + C_1\frac{s'd\log(p/\eta)}{N/T} + C_2h^2 + C_3\frac{R_Y^2 n_0 T s'\log(Tp/\eta)\log(5/\delta)}{N^2\epsilon^2}.
\end{aligned}
$$

Then by induction, it holds that

$$
\begin{aligned}
&\|\boldsymbol{\beta}^{T} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}^2 \\
\leq & \left(1 - \frac{2s_0}{s'}\xi - \frac{9s' - 10s_0}{9s'}\frac{\xi^2\underline{\gamma}}{2\overline{\gamma}} + \frac{10s_0}{9s'}\frac{\overline{\gamma}}{\underline{\gamma}}\right)^T \|\boldsymbol{\beta}^{0} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}^2 \\
&+ C_0 s'\sum_{t=0}^{T-1}\left(1 - \frac{2s_0}{s'}\xi - \frac{9s' - 10s_0}{9s'}\frac{\xi^2\underline{\gamma}}{2\overline{\gamma}} + \frac{10s_0}{9s'}\frac{\overline{\gamma}}{\underline{\gamma}}\right)^{T-t-1} \|\boldsymbol{\Delta}^t\|_\infty^2 \\
&+ C_1\frac{s'd\log(p/\eta)}{N/T} + C_2h^2 + C_3\frac{R_Y^2 n_0 T s'\log(Tp/\eta)\log(5/\delta)}{N^2\epsilon^2} \\
\leq & \left(1 - \frac{2s_0}{s'}\xi - \frac{9s' - 10s_0}{9s'}\frac{\xi^2\underline{\gamma}}{2\overline{\gamma}} + \frac{10s_0}{9s'}\frac{\overline{\gamma}}{\underline{\gamma}}\right)^T \|\boldsymbol{\beta}^{0} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}^2 \\
&+ C_0\frac{s'T^2R_d^2\log(1/\delta)\log(Tp/\eta)}{N^2\epsilon^2}\sum_{t=0}^{T-1}\left(1 - \frac{2s_0}{s'}\xi - \frac{9s' - 10s_0}{9s'}\frac{\xi^2\underline{\gamma}}{2\overline{\gamma}} + \frac{10s_0}{9s'}\frac{\overline{\gamma}}{\underline{\gamma}}\right)^{T-t-1} \sum_{k\in\{0\}\cup\mathcal{A}} R_k^2 \\
&+ C_1\frac{s'd\log(p/\eta)}{N/T} + C_2h^2 + C_3\frac{R_Y^2 n_0 T s'\log(Tp/\eta)\log(5/\delta)}{N^2\epsilon^2} \\
\leq & \left(1 - \frac{2s_0}{s'}\xi - \frac{9s' - 10s_0}{9s'}\frac{\xi^2\underline{\gamma}}{2\overline{\gamma}} + \frac{10s_0}{9s'}\frac{\overline{\gamma}}{\underline{\gamma}}\right)^T \|\boldsymbol{\beta}^{0} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}^2 \\
&+ C_0\frac{|\mathcal{A}|ps_0T^2\log(1/\delta)\log^2(N/\eta)\log(Tp/\eta)}{N^2\epsilon^2}\left(1 \vee \|\boldsymbol{\beta}^{0} - \boldsymbol{\beta}^{(0)}\|_{\boldsymbol{\Sigma}_{\mathcal{A}}}^2\right) \\
&+ C_1\frac{s'd\log(p/\eta)}{N/T} + C_2h^2 + C_3\frac{R_Y^2 n_0 T s'\log(Tp/\eta)\log(5/\delta)}{N^2\epsilon^2},
\end{aligned}
$$

where the last inequality follows from $s' \asymp s_0$, $R_d^2 \lesssim p\log(N/\eta)$, the choice of R_k, and the fact that

$$\sum_{t=0}^{T-1}\left(1-\frac{2s_0}{s'}\xi-\frac{9s'-10s_0}{9s'}\frac{\xi^2\underline{\gamma}}{2\overline{\gamma}}+\frac{10s_0}{9s'}\frac{\overline{\gamma}}{\underline{\gamma}}\right)^{T-t-1}\lesssim 1.$$

Therefore, conditioned on $\cap_{i=1}^4\mathcal{E}_i$, since $s' \asymp s_0$, we obtain

$$\begin{aligned}\|\boldsymbol{\beta}^T-\boldsymbol{\beta}^{(0)}\|_2 \leq&\sqrt{\frac{\overline{\gamma}}{\underline{\gamma}}}\left(1-\frac{2s_0}{s'}\xi-\frac{9s'-10s_0}{9s'}\frac{\xi^2\underline{\gamma}}{2\overline{\gamma}}+\frac{10s_0}{9s'}\frac{\overline{\gamma}}{\underline{\gamma}}\right)^{T/2}\|\boldsymbol{\beta}^0-\boldsymbol{\beta}^{(0)}\|_2\\ &+C_0\frac{T\sqrt{|\mathcal{A}|ps_0\log(1/\delta)\log^2(N/\eta)\log(Tp/\eta)}}{N\epsilon}\left(1\vee\|\boldsymbol{\beta}^0-\boldsymbol{\beta}^{(0)}\|_2\right)\\ &+C_1\sqrt{\frac{s_0 d\log(p/\eta)}{N/T}}+C_2h+C_3\frac{R_Y\sqrt{n_0Ts_0\log(Tp/\eta)\log(5/\delta)}}{N\epsilon}.\end{aligned}$$

Since $\xi = 1-0.296/L_0^4$, it can be deduced that

$$\begin{aligned}&1-\frac{9s'-10s_0}{9s'}\frac{9\xi^2}{22L_0^2}+\frac{10s_0}{9s'}\frac{11}{9}L_0^2\\ =&1-\frac{10s_0}{9s'}\frac{9}{22L_0^2}\left[\left(\frac{9s'}{10s_0}-1\right)\xi^2+\frac{22}{5}L_0^2\xi-2\left(\frac{11}{9}\right)^2L_0^4\right]\\ \leq&1-\frac{s_0}{s'}\frac{5}{11L_0^2}\left[\left(\frac{9s'}{10s_0}+\frac{17}{5}\right)\xi^2-2\left(\frac{11}{9}\right)^2L_0^4\right]\\ \leq&1-\frac{s_0}{s'}\frac{5}{11L_0^2}\left[\left(\frac{10}{9}\frac{\xi}{1-\xi}+\frac{17}{5}\right)\xi^2-2\left(\frac{11}{9}\right)^2L_0^4\right]<1.\end{aligned}$$

Finally, according to the conditions in Theorem 1 that $T \asymp \log N$, $s' \asymp s_0$, and $\|\boldsymbol{\beta}^0 - \boldsymbol{\beta}^{(0)}\|_2 \lesssim C$, we conclude that

$$\|\boldsymbol{\beta}^T-\boldsymbol{\beta}^{(0)}\|_2 \lesssim\sqrt{\frac{s_0 d\log(p/\eta)\log N}{N}}+\frac{\sqrt{|\mathcal{A}|ps_0\log(1/\delta)}\log^{5/2}(Np/\eta)}{N\epsilon}+h$$

holds with probability at least $1-\eta$.

□

A4.3. Proof of Corollary 1

Proof. Given the parameters chosen by Corollary 1, the first term of (B) in Theorem 1 is the main term. The result can be obtained directly through calculation. □

A4.4. Proof of Theorem 2

Proof. Similar to the proof of Theorem 1, we first define an event that happens with high probability:

$$\begin{aligned}\mathcal{E}_1' =&\{\tilde{R} \geq |(\check{\boldsymbol{X}}_i^{(0)})^\top \hat{\boldsymbol{\Theta}}_j|, \forall i \in \mathcal{I}^{(0)[t]}, \forall t = 0, \dots, T-1\} \\ &\cap \{\tilde{R} \geq |\check{Y}_i^{(0)}|, \forall i \in [n_0]\} \cap \{\tilde{R} \geq |(\boldsymbol{X}_i^{(0)})^\top \hat{\boldsymbol{\beta}}|, \forall i \in [n_0]\}.\end{aligned}$$

We assume all the data are process on the target and thus omit some subscript and superscript for convenience, i.e., $n := n_0$, $\boldsymbol{\beta} := \boldsymbol{\beta}^{(0)}$, $\boldsymbol{X} := \boldsymbol{X}^{(0)}$, $\boldsymbol{Y} := \boldsymbol{Y}^{(0)}$, and $\boldsymbol{W} := \boldsymbol{W}^{(0)}$ across this proof.

Define the untruncated debiased estimator as

$$\hat{\beta}_j^{(db)} := \hat{\beta}_j + \frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top \check{\boldsymbol{X}}^\top \{\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{g}}(\boldsymbol{W})\}. \tag{16}$$

This is the untruncated version of (3.5). Denote

$$\hat{\boldsymbol{g}}(\boldsymbol{W}) := \{\hat{g}(\boldsymbol{W}_i)\}_{i=1}^n, \qquad \check{\boldsymbol{Y}} := \boldsymbol{Y} - \hat{\boldsymbol{g}}(\boldsymbol{W}),$$

and

$$\check{\boldsymbol{X}}_j := \boldsymbol{X}_j - \hat{\boldsymbol{f}}_j(\boldsymbol{W}), \qquad \tilde{\boldsymbol{X}}_j := \boldsymbol{X}_j - \boldsymbol{f}_j(\boldsymbol{W}), \qquad j = 1, \dots, p.$$

Throughout this proof, $\check{\boldsymbol{Y}}$ is used through the residual

$$\check{\boldsymbol{Y}} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{g}}(\boldsymbol{W}),$$

which corresponds to the second debiased estimator in the partially linear model literature. For notational simplicity, we write

$$\boldsymbol{f}(\boldsymbol{W}) := \{\boldsymbol{f}(\boldsymbol{W}_i)^\top\}_{i=1}^n \in \mathbb{R}^{n \times p}, \qquad \hat{\boldsymbol{f}}(\boldsymbol{W}) := \{\hat{\boldsymbol{f}}(\boldsymbol{W}_i)^\top\}_{i=1}^n \in \mathbb{R}^{n \times p}.$$

Then

$$\check{\boldsymbol{X}} = \boldsymbol{X} - \hat{\boldsymbol{f}}(\boldsymbol{W}), \qquad \tilde{\boldsymbol{X}} = \boldsymbol{X} - \boldsymbol{f}(\boldsymbol{W}), \qquad \check{\boldsymbol{X}} - \tilde{\boldsymbol{X}} = \boldsymbol{f}(\boldsymbol{W}) - \hat{\boldsymbol{f}}(\boldsymbol{W}).$$

By the model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{g}(\boldsymbol{W}) + \boldsymbol{\varepsilon},$$

we have

$$\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{g}}(\boldsymbol{W}) = \boldsymbol{\varepsilon} - \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \{\hat{\boldsymbol{g}}(\boldsymbol{W}) - \boldsymbol{g}(\boldsymbol{W})\}.$$

Moreover, since

$$\boldsymbol{X} = \check{\boldsymbol{X}} + \hat{\boldsymbol{f}}(\boldsymbol{W}) = \check{\boldsymbol{X}} + \boldsymbol{f}(\boldsymbol{W}) + \{\hat{\boldsymbol{f}}(\boldsymbol{W}) - \boldsymbol{f}(\boldsymbol{W})\} = \check{\boldsymbol{X}} + \boldsymbol{f}(\boldsymbol{W}) + \{\tilde{\boldsymbol{X}} - \check{\boldsymbol{X}}\},$$

it follows that

$$\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \check{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + \boldsymbol{f}(\boldsymbol{W})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + (\tilde{\boldsymbol{X}}-\check{\boldsymbol{X}})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}).$$

Consequently,

$$\begin{aligned}\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{g}}(\boldsymbol{W}) = \boldsymbol{\varepsilon} - \check{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) - \boldsymbol{f}(\boldsymbol{W})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\ - (\tilde{\boldsymbol{X}}-\check{\boldsymbol{X}})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) - \{\hat{\boldsymbol{g}}(\boldsymbol{W}) - \boldsymbol{g}(\boldsymbol{W})\}.\end{aligned}$$

Therefore,

$$\begin{aligned}\hat{\beta}_j^{(db)} - \beta_j &= \hat{\beta}_j - \beta_j + \frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top \check{\boldsymbol{X}}^\top\{\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{g}}(\boldsymbol{W})\} \\
&= \hat{\beta}_j - \beta_j + \frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top \check{\boldsymbol{X}}^\top\boldsymbol{\varepsilon} - \frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top \check{\boldsymbol{X}}^\top\check{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\
&\quad - \frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top \check{\boldsymbol{X}}^\top\boldsymbol{f}(\boldsymbol{W})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) - \frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top \check{\boldsymbol{X}}^\top(\tilde{\boldsymbol{X}}-\check{\boldsymbol{X}})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\
&\quad - \frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top \check{\boldsymbol{X}}^\top\{\hat{\boldsymbol{g}}(\boldsymbol{W}) - \boldsymbol{g}(\boldsymbol{W})\} \\
&= \frac{1}{n}\boldsymbol{\Theta}_j^\top\tilde{\boldsymbol{X}}^\top\boldsymbol{\varepsilon} + \underbrace{\frac{1}{n}(\hat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j)^\top\tilde{\boldsymbol{X}}^\top\boldsymbol{\varepsilon}}_{:=E_0} - \underbrace{\frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top(\tilde{\boldsymbol{X}}-\check{\boldsymbol{X}})^\top\boldsymbol{\varepsilon}}_{:=E_1} \\
&\quad + \underbrace{\left(\mathrm{e}_j - \frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}}\right)^\top(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}_{:=E_2} + \underbrace{\frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top\tilde{\boldsymbol{X}}^\top(\check{\boldsymbol{X}}-\tilde{\boldsymbol{X}})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}_{:=E_3} \\
&\quad - \underbrace{\frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top\tilde{\boldsymbol{X}}^\top\boldsymbol{f}(\boldsymbol{W})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}_{:=E_4} - \underbrace{\frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top(\check{\boldsymbol{X}}-\tilde{\boldsymbol{X}})^\top\boldsymbol{f}(\boldsymbol{W})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}_{:=E_5} \\
&\quad - \underbrace{\frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top\check{\boldsymbol{X}}^\top\{\hat{\boldsymbol{g}}(\boldsymbol{W}) - \boldsymbol{g}(\boldsymbol{W})\}}_{:=E_6}.\end{aligned}$$

By Hölder's inequality, it can be obtained that,

$$E_0 \le \|\hat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\|_1\|\tfrac{1}{n}\tilde{\boldsymbol{X}}^\top\boldsymbol{\varepsilon}\|_\infty, \tag{17}$$

$$E_1 \le \|\hat{\boldsymbol{\Theta}}_j\|_1\|\tfrac{1}{n}(\tilde{\boldsymbol{X}}-\check{\boldsymbol{X}})^\top\boldsymbol{\varepsilon}\|_\infty, \tag{18}$$

$$E_2 \le \|\mathrm{e}_j - \tfrac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}}\|_\infty\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1, \tag{19}$$

$$E_3 \le \|\hat{\boldsymbol{\Theta}}_j\|_1\|\tfrac{1}{n}\tilde{\boldsymbol{X}}^\top(\check{\boldsymbol{X}}-\tilde{\boldsymbol{X}})\|_\infty\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1, \tag{20}$$

$$E_4 \le \|\hat{\boldsymbol{\Theta}}_j\|_1\|\tfrac{1}{n}\tilde{\boldsymbol{X}}^\top\boldsymbol{f}(\boldsymbol{W})\|_\infty\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1, \tag{21}$$

$$E_5 \le \|\hat{\boldsymbol{\Theta}}_j\|_1\max_{j=1,\ldots,p}\|\hat{\boldsymbol{f}}_j(\boldsymbol{W}) - \boldsymbol{f}_j(\boldsymbol{W})\|_n\|\boldsymbol{f}(\boldsymbol{W})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|_n, \tag{22}$$

and

$$E_6 \leq \|\hat{\mathbf{\Theta}}_j\|_1 \|\tfrac{1}{n}\check{\boldsymbol{X}}^\top(\hat{\boldsymbol{g}}(\boldsymbol{W}) - \boldsymbol{g}(\boldsymbol{W}))\|_\infty. \tag{23}$$

By Corollary 1 and the fact that $0 < L_0^{-1} \leq \lambda_{\min}(\mathbf{\Sigma}) \leq \lambda_{\max}(\mathbf{\Sigma}) \leq L_0 < \infty$ in Assumption 1, it holds that

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 = O_p(s_0\sqrt{\frac{\log p}{n}}), \tag{24}$$

and

$$\|\boldsymbol{f}(\boldsymbol{W})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_n = \frac{1}{\sqrt{n}}\|\boldsymbol{f}(\boldsymbol{W})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_2 = O_p(\sqrt{\frac{s_0 \log p}{n}}). \tag{25}$$

Also according to Assumption 1, standard tail bounds yield

$$\|\frac{1}{n}\tilde{\boldsymbol{X}}^\top\boldsymbol{\varepsilon}\|_\infty = O_p(\sqrt{\frac{\log p}{n}}), \tag{26}$$

and

$$\|\frac{1}{n}\tilde{\boldsymbol{X}}^\top\boldsymbol{f}(\boldsymbol{W})\|_\infty = O_p(\sqrt{\frac{\log p}{n}}), \tag{27}$$

where we have used the fact that $f_j(\boldsymbol{W}) = \boldsymbol{X}_j - \tilde{\boldsymbol{X}}_j$ is sub-Gaussian that is implied by Assumption 1(ii).

Under Assumptions 1 and 3 , according to the Lemma 1, for any $t_n \geq \tilde{r}_n$,

$$\|\hat{\boldsymbol{f}}_j - \boldsymbol{f}_j\|_n = \sqrt{\frac{1}{n}\sum_i(\hat{f}_j(\boldsymbol{W}_i) - f_j(\boldsymbol{W}_i))^2} \leq C\sqrt{d}\,t_n \tag{28}$$

holds with probability at least $1 - c\exp(-Cnt_n^2)$. Due to the union bound, it can be deduced that the following rsult

$$\max_{j\in[p]}\|\hat{\boldsymbol{f}}_j - \boldsymbol{f}_j\|_n \leq C\sqrt{d}\,t_n \tag{29}$$

with probability at least $1 - c\exp(\log p - Cnt_n^2)$. Moreover, based on Corollary 1 and Lemma 1, we have

$$\|\hat{\boldsymbol{g}} - \boldsymbol{g}\|_n = \sqrt{\frac{1}{n}\sum_i(\hat{g}(\boldsymbol{W}_i) - g(\boldsymbol{W}_i))^2} = O_p(\sqrt{s_0}\tilde{r}_n \vee \sqrt{\frac{s_0 \log p}{n}}). \tag{30}$$

By Cauchy-Schwarz's inequality, it can be obtained that

$$|\frac{1}{n}(\tilde{\boldsymbol{X}}_j - \check{\boldsymbol{X}}_j)^\top \boldsymbol{\varepsilon}| = |\frac{1}{n}[\hat{\boldsymbol{f}}_j - \boldsymbol{f}_j]^\top \boldsymbol{\varepsilon}|, \tag{31}$$

$$|\frac{1}{n}\check{\boldsymbol{X}}_j^\top (\check{\boldsymbol{X}}_{j'} - \tilde{\boldsymbol{X}}_{j'})| \leq \|\hat{\boldsymbol{f}}_j - \boldsymbol{f}_j\|_n \|\hat{\boldsymbol{f}}_{j'} - \boldsymbol{f}_{j'}\|_n + |\frac{1}{n}\tilde{\boldsymbol{X}}_j^\top [\hat{\boldsymbol{f}}_{j'} - \boldsymbol{f}_{j'}]|, \tag{32}$$

and

$$|\frac{1}{n}\check{\boldsymbol{X}}_j^\top (\hat{\boldsymbol{g}} - \boldsymbol{g})| \leq \|\hat{\boldsymbol{f}}_j - \boldsymbol{f}_j\|_n \|\hat{\boldsymbol{g}} - \boldsymbol{g}\|_n + |\frac{1}{n}\tilde{\boldsymbol{X}}_j^\top (\hat{\boldsymbol{g}} - \boldsymbol{g})|, \tag{33}$$

where we use $\boldsymbol{f}_j$, $\hat{\boldsymbol{f}}_j$, $\boldsymbol{g}$, and $\hat{\boldsymbol{g}}$ as the abbreviations for $\boldsymbol{f}_j(\boldsymbol{W})$, $\hat{\boldsymbol{f}}_j(\boldsymbol{W})$, $\boldsymbol{g}(\boldsymbol{W})$, and $\hat{\boldsymbol{g}}(\boldsymbol{W})$, respectively. Additionally, note that $\mathbb{E}(\varepsilon_i \mid \boldsymbol{W}_i) = 0$ and $\mathbb{E}(\tilde{Y}_i \mid \boldsymbol{W}_i) = \mathbb{E}(\tilde{\boldsymbol{X}}_{ij} \mid \boldsymbol{W}_i) = 0$ (by construction of $\tilde{Y}_i$ and $\tilde{\boldsymbol{X}}_{ij}, i = 1, \ldots, n, j = 1, \ldots, p$). The remaining argument is derived based on the results from empirical process theory and local function complexity. In particular, under the independent sampling assumption, joint with Lemma 1 and (28)-(30), from (31)-(33), it can be yielded that

$$\max_{j=1,\ldots,p} |\frac{1}{n}[\boldsymbol{f}_j(\boldsymbol{W}) - \hat{\boldsymbol{f}}_j(\boldsymbol{W})]^\top \boldsymbol{\varepsilon}| = O_p(t_n^2), \tag{34}$$

$$\max_{j,j' \in \{1,\ldots,p\}} |\frac{1}{n}\tilde{\boldsymbol{X}}_j^\top [\hat{\boldsymbol{f}}_{j'}(\boldsymbol{W}) - \boldsymbol{f}_{j'}(\boldsymbol{W})]| = O_p(t_n^2), \tag{35}$$

and

$$\max_{j=1,\ldots,p} |\frac{1}{n}\tilde{\boldsymbol{X}}_j^\top (\hat{\boldsymbol{g}}(\boldsymbol{W}) - \boldsymbol{g}(\boldsymbol{W}))| = O_p(s_0 t_n^2) + O_p(\frac{s_0 \log p}{n}), \tag{36}$$

provided with $t_n \geq r_n$ and $nt_n^2 \gtrsim \log p$. In our analysis, it suffices to choose $t_n = (\tilde{r}_n \vee \sqrt{\frac{\log p}{n}})$. Consequently, we have

$$\|\frac{1}{n}(\tilde{\boldsymbol{X}} - \check{\boldsymbol{X}})^\top \boldsymbol{\varepsilon}\|_\infty = O_p(\tilde{r}_n^2 \vee \frac{\log p}{n}), \tag{37}$$

$$\|\frac{1}{n}\check{\boldsymbol{X}}^\top (\check{\boldsymbol{X}} - \tilde{\boldsymbol{X}})\|_\infty = O_p(\tilde{r}_n^2 \vee \frac{\log p}{n}), \tag{38}$$

and

$$\|\frac{1}{n}\check{\boldsymbol{X}}^\top (\hat{\boldsymbol{g}}(\boldsymbol{W}) - \boldsymbol{g}(\boldsymbol{W}))\|_\infty = O_p(s_0 \tilde{r}_n^2 \vee \frac{s_0 \log p}{n}). \tag{39}$$

Next, we establish the bounds related to $\mathbf{\Theta}_j$. Although Algorithm 2 is implemented with the adjusted covariates $\check{\boldsymbol{X}}$, the empirical process bounds established above imply that

$$\left\|\frac{1}{n}\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}} - \frac{1}{n}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}}\right\|_\infty = O_p\left(\tilde{r}_n^2 \vee \frac{\log p}{n}\right),$$

which is asymptotically negligible under Assumption 4. Therefore, the argument of Lemma 9 in Cai et al. (2026) can be applied to the adjusted design $\check{\boldsymbol{X}}$, yielding

$$\|\hat{\mathbf{\Theta}}_j - \mathbf{\Theta}_j\|_1 \lesssim \tilde{s}_j\sqrt{\frac{\log p \log n}{n}} + \frac{\tilde{s}_j^{3/2}\log p \log^3 n\sqrt{\log(1/\delta)}}{n\epsilon} + \tilde{s}_j^2\left(\tilde{r}_n^2 \vee \frac{\log p}{n}\right),$$

with probability at least $1-\exp\{-C\log(n)\}$. Under Assumption 4, the last term is negligible. Given that

$$\sqrt{\frac{\bar{s}\log p\log(1/\delta)}{n}}\log^{3/2} n \lesssim \epsilon$$

as Theorem 2 assumes, it follows that

$$\|\hat{\mathbf{\Theta}}_j - \mathbf{\Theta}_j\|_1 = O_p\left(\tilde{s}_j\sqrt{\frac{\log p\log n}{n}}\right), \tag{40}$$

and

$$\|\hat{\mathbf{\Theta}}_j\|_1 = O_p(\tilde{s}_j). \tag{41}$$

Therefore, $\|\mathbf{e}_j - \frac{1}{n}\hat{\mathbf{\Theta}}_j^\top\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}}\|_\infty$ can be bounded as follows:

$$\begin{aligned}
&\left\|\frac{1}{n}\hat{\mathbf{\Theta}}_j^\top\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}} - \mathbf{e}_j\right\|_\infty\\
\le&\left\|\hat{\mathbf{\Theta}}_j - \mathbf{\Theta}_j\right\|_1\left\|\frac{1}{n}\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}}\right\|_\infty + \|\mathbf{\Theta}_j\|_1\left\|\frac{1}{n}\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}} - \mathbb{E}(\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}_i^\top)\right\|_\infty\\
\le&\left\|\hat{\mathbf{\Theta}}_j - \mathbf{\Theta}_j\right\|_1\left\|\frac{1}{n}\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}}\right\|_\infty + \|\mathbf{\Theta}_j\|_1\left(\left\|\frac{1}{n}\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}} - \frac{1}{n}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}}\right\|_\infty + \left\|\frac{1}{n}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} - \mathbb{E}(\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}_i^\top)\right\|_\infty\right).
\end{aligned}$$

According to (38), it can be shown that

$$\begin{aligned}
&\|\frac{1}{n}\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}} - \frac{1}{n}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}}\|_\infty\\
\le&\left\|\frac{1}{n}\tilde{\boldsymbol{X}}^\top(\check{\boldsymbol{X}}-\tilde{\boldsymbol{X}})\right\|_\infty + \left\|\frac{1}{n}(\check{\boldsymbol{X}}-\tilde{\boldsymbol{X}})^\top\tilde{\boldsymbol{X}}\right\|_\infty + \left\|\frac{1}{n}(\check{\boldsymbol{X}}-\tilde{\boldsymbol{X}})^\top(\check{\boldsymbol{X}}-\tilde{\boldsymbol{X}})\right\|_\infty\\
=&O_p\left(r_n^2 \vee \frac{\log p}{n}\right).
\end{aligned}$$

Moreover, by standard tail bounds for sub-exponential random variables, we have

$$\left\|\frac{1}{n}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} - \mathbb{E}\left(\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}_i^\top\right)\right\|_\infty = O_p\left(\sqrt{\frac{\log p}{n}}\right).$$

Consequently,

$$\left\|\frac{1}{n}\hat{\boldsymbol{\Theta}}_j^\top\check{\boldsymbol{X}}^\top\check{\boldsymbol{X}} - \mathbf{e}_j\right\|_\infty = O_p(\tilde{s}_j\sqrt{\frac{\log p}{n}} + \tilde{s}_j^{3/2}(\tilde{r}_n^2 \vee \frac{\log p}{n})). \tag{42}$$

Under Assumption 4, putting all the pieces above together, we obtain the following probabilistic bounds for $E_0, \ldots, E_6$. In each case, we upper bound the right hand side of (17)-(23) by the intermediate estimates cited. By inserting (3.7)-(3.9) in Assumption 4 into (17)-(23), the rate of orders are obtained as per:

$$E_0 \underset{(40),(26)}{=} O_p\Big(\tilde{s}_j\frac{\log p\sqrt{\log n}}{n}\Big) = o_p(n^{-1/2}),$$

$$E_1 \underset{(41),(26)}{=} O_p\Big(\tilde{s}_j\Big(1+\sqrt{\tfrac{\log p}{n}}\Big)\left(\tilde{r}_n^2 \vee \tfrac{\log p}{n}\right)\Big) = o_p(n^{-1/2}),$$

$$E_2 \underset{(42),(24)}{=} O_p\Big(s_0\tilde{s}_j\frac{\log p}{n} + s_0\tilde{s}_j^{3/2}\left(\tilde{r}_n^2 \vee \tfrac{\log p}{n}\right)\sqrt{\tfrac{\log p}{n}}\Big) = o_p(n^{-1/2}),$$

$$E_3 \underset{(41),(38),(24)}{=} O_p\Big(s_0\tilde{s}_j\Big(\tfrac{\log p}{n} + \sqrt{\tfrac{\log p}{n}}\Big)\left(\tilde{r}_n^2 \vee \tfrac{\log p}{n}\right)\Big) = o_p(n^{-1/2}),$$

$$E_4 \underset{(41),(27),(24)}{=} O_p\Big(s_0\tilde{s}_j\Big(\tfrac{\log p}{n} + \sqrt{\tfrac{\log p}{n}}\Big)\Big) = o_p(n^{-1/2}),$$

$$E_5 \underset{(41),(29),(25)}{=} O_p\Big(\sqrt{s_0}\,\tilde{s}_j\Big(\tfrac{\log p}{n} + \sqrt{\tfrac{\log p}{n}}\Big)\left(\tilde{r}_n^2 \vee \tfrac{\log p}{n}\right)\Big) = o_p(n^{-1/2}),$$

and

$$E_6 \underset{(41),(33)}{=} = O_p\Big(s_0\tilde{s}_j\Big(1+\sqrt{\tfrac{\log p}{n}}\Big)\frac{\log p}{n}\Big) = o_p(n^{-1/2}).$$

Thus, by central limit theorem,

$$\frac{\sqrt{n}(\hat{\beta}_j^{(db)} - \beta_j)}{V_j} \xrightarrow{D} \mathrm{N}(0,1),$$

where $V_j^2 = \mathbf{\Theta}_j \mathbb{E}(\tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{X}}_i^\top)\mathbf{\Theta}_j^\top \sigma^2 = \Theta_{j,j}\sigma^2$ with $\Theta_{j,j}$ being the (j,j)-th element of $\mathbf{\Theta}$. On the event $\mathcal{E}_1'$, the truncation in (3.5) is inactive, and hence $\hat{\beta}_j^{(de)} = \hat{\beta}_j^{(db)}$. We have

$$\sqrt{n_0}(\hat{\beta}_j^{(dp)} - \beta_j) = u_j + v_j,$$

where $u_j \xrightarrow{D} \mathrm{N}(0, V_j^2)$ and $v_j = O_p\{s_0 \log(p)\log(n_0)/\sqrt{n_0}\}$.

Finally, notice that $\sqrt{n}\sup|\tilde{\Delta}_j| \leq C\sqrt{\log n/n}\log(1/\delta)/\epsilon$ for some constant C with probability approaching 1. It implies that $|\tilde{\Delta}_j| = o_p(n^{-1/2})$. Thus, $\hat{\boldsymbol{\beta}}^{(dp)}$ has the same asymptotic distribution as $\hat{\boldsymbol{\beta}}^{(de)}$. It remains to verify the privacy property for the debiased term. The sensitivity of the debiased term $\Pi_{\tilde{R}}(\hat{\mathbf{\Theta}}_j^\top \check{\boldsymbol{X}}_i)(\Pi_{\tilde{R}}(\check{Y}_i) - \Pi_{\tilde{R}}(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}))/n$ satisfies

$$\begin{aligned}
&\sup_{(\boldsymbol{X}_i, \check{\boldsymbol{X}}_i, \check{Y}_i), (\boldsymbol{X}_i', \check{\boldsymbol{X}}_i', \check{Y}_i')} \frac{1}{n}\Big| \Pi_{\tilde{R}}(\hat{\mathbf{\Theta}}_j^\top \check{\boldsymbol{X}}_i)\{\Pi_{\tilde{R}}(\check{Y}_i) - \Pi_{\tilde{R}}(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}})\} \\
&\qquad - \Pi_{\tilde{R}}(\hat{\mathbf{\Theta}}_j^\top \check{\boldsymbol{X}}_i')\{\Pi_{\tilde{R}}(\check{Y}_i') - \Pi_{\tilde{R}}((\boldsymbol{X}_i')^\top \hat{\boldsymbol{\beta}})\}\Big| \\
&\leq \frac{2}{n}(\tilde{R}^2 + \tilde{R}^2) = \frac{4\tilde{R}^2}{n}.
\end{aligned}$$

Therefore, the noise added in the debiased term satisfies the Gaussian mechanism. According to the composition property, the calculation of $\hat{\beta}_j^{(dp)}$ is (ϵ, δ)-STDP. □

A4.5. Proof of Corollary 2

Proof. Given the results obtained by Theorem 2, the remaining task in this proof is to show: (i) the consistency of $\hat{V}_j^{(dp)}$; (ii) the differential privacy property of $\hat{V}_j^{(dp)}$.

(i)

For the variance estimation, we have

$$\begin{aligned}
\hat{\sigma}^2 - \sigma^2 &= \frac{1}{n}\sum_{i=1}^n (\Pi_{\tilde{R}}(\check{Y}_i) - \Pi_{\tilde{R}}(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}))^2 + \Delta_\sigma - \sigma^2 \\
&= \frac{1}{n}\sum_{i=1}^n (\Pi_{\tilde{R}}(\check{Y}_i) - \boldsymbol{X}_i^\top \boldsymbol{\beta})^2 - \sigma^2 \\
&+ \frac{1}{n}\sum_{i=1}^n (\Pi_{\tilde{R}}(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}) - \boldsymbol{X}_i^\top \boldsymbol{\beta})^2 - \frac{2}{n}\sum_{i=1}^n (\Pi_{\tilde{R}}(\check{Y}_i) - \boldsymbol{X}_i^\top \boldsymbol{\beta})(\Pi_{\tilde{R}}(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}) - \boldsymbol{X}_i^\top \boldsymbol{\beta}) + \Delta_\sigma \\
&= o_p(1).
\end{aligned}$$

According to the condition in Theorem 2 that $\sqrt{\frac{\bar{s}\log p \log(1/\delta)}{n}}\log^{3/2} n \lesssim \epsilon$ and Assumption 4, we have

$$\|\hat{\mathbf{\Theta}}_j - \mathbf{\Theta}_j\|_2 = O_p(\sqrt{\frac{\tilde{s}_j \log p \log n}{n}}) = o_p(1).$$

Hence, $\hat{V}_j^{(dp)}$ is a consistent estimation of V_j. According to the Slutsky's theorem the asymptotic normality for $\sqrt{n}(\hat{\beta}_j^{(dp)} - \beta_j)/\hat{V}_j^{(dp)}$ holds.

(ii)

In Theorem 2, we have shown the privacy property of $\hat{\beta}_j^{(dp)}$. Moreover, according to Lemma 9 in Cai et al. (2026), the estimation $\hat{\boldsymbol{\Theta}}_j$ is also privacy guaranteed. It remains to verify the privacy property for the estimation $\hat{\sigma}$.

The sensitivity of the sample variance $\frac{1}{n}\sum_{i=1}^n (\Pi_{\tilde{R}}(\check{Y}_i) - \Pi_{\tilde{R}}(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}))^2$ satisfies

$$\sup_{(\boldsymbol{X}_i, \check{\boldsymbol{X}}_i, \check{Y}_i), (\boldsymbol{X}_i', \check{\boldsymbol{X}}_i', \check{Y}_i')} \frac{1}{n} |(\Pi_{\tilde{R}}(\check{Y}_i) - \Pi_{\tilde{R}}(\boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}))^2 - (\Pi_{\tilde{R}}(\check{Y}_i') - \Pi_{\tilde{R}}((\boldsymbol{X}_i')^\top \hat{\boldsymbol{\beta}}))^2| \leq \frac{2}{n}(2\tilde{R})^2.$$

Therefore, the noise added in satisfies the Gaussian mechanism. According to the composition property, the privatized inference is (ϵ, δ)-STDP.

□

A4.6. Derivation of e-value

This subsection verifies that the statistic used in the proposed multiple testing procedure is an asymptotic e-value under the null hypotheses.

Lemma 3. We use the notation in the proof of Theorem 2, and define

$$q_n(z) = \sqrt{\frac{\log p + z}{n}}, \qquad a_n(z) = \left(\tilde{r}_n^2 \vee \frac{\log p + z}{n}\right).$$

Under Assumptions 1, 2, and 3, together with the tuning conditions in Theorem 2, there exist constants $C, c > 0$ such that, for every $z \geq 1$, with probability at least $1 - Ce^{-cz}$,

$$\max_{1 \leq \ell \leq p} \|\hat{\boldsymbol{f}}_\ell(\boldsymbol{W}) - \boldsymbol{f}_\ell(\boldsymbol{W})\|_n \leq C a_n(z)^{1/2}, \tag{43}$$

$$\|\hat{\boldsymbol{g}}(\boldsymbol{W}) - \boldsymbol{g}(\boldsymbol{W})\|_n \leq C\sqrt{s_0}\, a_n(z)^{1/2}. \tag{44}$$

Proof. The first bound follows by applying the RKHS least-squares tail bound in Lemmas S.1–S.2 of Zhu (2017) to each regression of X_ℓ on $\boldsymbol{W}$, and then taking a union bound over $\ell \in [p]$.

For the second bound, view the pseudo-response $Y_i - \boldsymbol{X}_i^\top \hat{\boldsymbol{\beta}}$ as having conditional mean

$$g(\boldsymbol{W}_i) - f(\boldsymbol{W}_i)^\top (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Lemma S.10 of Zhu (2017) gives the KRR error around this conditional mean at rate $a_n(z)^{1/2}$. The remaining bias term is bounded by the standard sparse-direction empirical bound for $\boldsymbol{f}(\boldsymbol{W})$, together with $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 \leq C\sqrt{s_0} q_n(z)$. Since $a_n(z) \geq q_n(z)^2$, this yields (44). □

Lemma 4. Let the notation of Lemma 3 and proof of Theorem 2 be used. For $z \geq 1$, define

$$r_{\Theta,j}(z) = \tilde{s}_j \sqrt{\frac{(\log p + z)\log n}{n}} + \frac{\tilde{s}_j^{3/2}(\log p + z)\log^3 n \sqrt{\log(1/\delta)}}{n\epsilon} + \tilde{s}_j^2 a_n(z),$$
$$u_{\Theta,j}(z) = r_{\Theta,j}(z) + \tilde{s}_j q_n(z) + \tilde{s}_j^{3/2} a_n(z).$$

Under the same conditions in Theorem 2, there exist constants $C, c > 0$, independent of n, p, z, j, such that, for every $z \geq 1$, with probability at least $1 - Ce^{-cz}$, the following bounds hold:

$$|E_0| \leq C r_{\Theta,j}(z) q_n(z), \qquad |E_1| \leq C \tilde{s}_j a_n(z),$$
$$|E_2| \leq C s_0 q_n(z) u_{\Theta,j}(z), \quad |E_3| \leq C s_0 \tilde{s}_j a_n(z) q_n(z),$$
$$|E_4| \leq C s_0 \tilde{s}_j q_n(z)^2, \qquad |E_5| \leq C\sqrt{s_0}\, \tilde{s}_j a_n(z),$$
$$|E_6| \leq C s_0 \tilde{s}_j a_n(z).$$

Proof. We first collect the preliminary bounds needed below. Throughout the proof, $C, c > 0$ denote constants independent of n, p, z, and j, whose values may change from line to line.

By Corollary 1, the transfer and privacy-rate conditions in Theorem 2, and Assumptions 1–4, the preliminary estimator satisfies

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \leq C s_0 q_n(z), \qquad \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 \leq C\sqrt{s_0}\, q_n(z) \tag{45}$$

with probability at least $1 - Ce^{-cz}$. Moreover, $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is supported on a set of size $O(s_0)$. Therefore, the sparse-direction empirical bound for $\boldsymbol{f}(\boldsymbol{W})$ gives

$$\begin{aligned} \|\boldsymbol{f}(\boldsymbol{W})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_n &\leq \sup_{\substack{\boldsymbol{v} \neq 0 \\ \|\boldsymbol{v}\|_0 \leq Cs_0}} \frac{\|\boldsymbol{f}(\boldsymbol{W})\boldsymbol{v}\|_n}{\|\boldsymbol{v}\|_2} \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 \\ &\leq C\sqrt{s_0}\, q_n(z). \end{aligned} \tag{46}$$

In addition, the localized multiplier and covariance bounds following from (43)–(44) yield

$$\left\| \frac{1}{n}(\tilde{\boldsymbol{X}} - \check{\boldsymbol{X}})^\top \boldsymbol{\varepsilon} \right\|_\infty \leq C a_n(z), \tag{47}$$
$$\left\| \frac{1}{n}\check{\boldsymbol{X}}^\top (\check{\boldsymbol{X}} - \tilde{\boldsymbol{X}}) \right\|_\infty \leq C a_n(z), \tag{48}$$
$$\left\| \hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} - \frac{1}{n}\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}} \right\|_\infty \leq C a_n(z), \tag{49}$$
$$\left\| \frac{1}{n}\check{\boldsymbol{X}}^\top \{\hat{\boldsymbol{g}}(\boldsymbol{W}) - \boldsymbol{g}(\boldsymbol{W})\} \right\|_\infty \leq C s_0 a_n(z), \tag{50}$$

with probability at least $1 - Ce^{-cz}$, where we define

$$\hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} := \frac{1}{n}\check{\boldsymbol{X}}^\top \check{\boldsymbol{X}}, \quad \boldsymbol{\Sigma}_{\tilde{\boldsymbol{X}}} := \mathbb{E}(\tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{X}}_i^\top), \quad \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{X}}} := \frac{1}{n}\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}}.$$

To account for the use of $\check{\boldsymbol{X}}$, observe that the additional gradient perturbation at $\boldsymbol{\Theta}_j$ is bounded by

$$\left\| \left(\hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} - \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{X}}} \right) \boldsymbol{\Theta}_j \right\|_\infty \le \left\| \hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} - \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{X}}} \right\|_\infty \|\boldsymbol{\Theta}_j\|_1.$$

Assumption 4 implies

$$\|\boldsymbol{\Theta}_j\|_1 \le \sqrt{\tilde{s}_j}\, \|\boldsymbol{\Theta}_j\|_2 \le C\sqrt{\tilde{s}_j} \le C\tilde{s}_j.$$

Consequently,

$$\left\| \left(\hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} - \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{X}}} \right) \boldsymbol{\Theta}_j \right\|_\infty \le C\tilde{s}_j a_n(z).$$

Following the same proof as in Lemma 9 of Cai et al. (2026) but replacing $\tilde{\boldsymbol{X}}$ with $\check{\boldsymbol{X}}$ contributes at most $C\tilde{s}_j^2 a_n(z)$ to the ℓ_1 error. It follows that

$$\|\hat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\|_1 \le C r_{\Theta,j}(z) \tag{51}$$

with probability at least $1 - Ce^{-cz}$. By the triangle inequality,

$$\|\hat{\boldsymbol{\Theta}}_j\|_1 \le \|\boldsymbol{\Theta}_j\|_1 + \|\hat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\|_1 \le C\sqrt{\tilde{s}_j} + C r_{\Theta,j}(z) \le C\tilde{s}_j, \tag{52}$$

where the last inequality follows from the rate conditions in Theorem 2.

Under Assumption 1, each product $\tilde{X}_{i\ell}\varepsilon_i$ is sub-exponential. Therefore, Bernstein's inequality and a union bound over $\ell \in [p]$ give

$$\left\| \frac{1}{n}\tilde{\boldsymbol{X}}^\top \boldsymbol{\varepsilon} \right\|_\infty \le C q_n(z) \tag{53}$$

with probability at least $1 - Ce^{-cz}$. Similarly, each $\tilde{X}_{i\ell} f_m(\boldsymbol{W}_i)$ is sub-exponential, and a union bound over $(\ell, m) \in [p]^2$ gives

$$\left\| \frac{1}{n}\tilde{\boldsymbol{X}}^\top \boldsymbol{f}(\boldsymbol{W}) \right\|_\infty \le C q_n(z). \tag{54}$$

Finally, we establish the approximate inverse-score bound. Since

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{X}}} \boldsymbol{\Theta}_j = \mathbf{e}_j,$$

we have

$$\mathbf{e}_j^\top - \hat{\boldsymbol{\Theta}}_j^\top \hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} = (\boldsymbol{\Theta}_j - \hat{\boldsymbol{\Theta}}_j)^\top \hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} + \boldsymbol{\Theta}_j^\top \left(\boldsymbol{\Sigma}_{\tilde{\boldsymbol{X}}} - \hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} \right).$$

Hence,

$$\left\| \mathbf{e}_j^\top - \hat{\boldsymbol{\Theta}}_j^\top \hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} \right\|_\infty \leq \|\hat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\|_1 \|\hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}}\|_\infty + \|\boldsymbol{\Theta}_j\|_1 \left\| \hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} - \boldsymbol{\Sigma}_{\tilde{\boldsymbol{X}}} \right\|_\infty .$$

By Bernstein's inequality, $\left\| \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{X}}} - \boldsymbol{\Sigma}_{\tilde{\boldsymbol{X}}} \right\|_\infty \leq C q_n(z)$. Combining this with the covariance perturbation bound gives

$$\left\| \hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} - \boldsymbol{\Sigma}_{\tilde{\boldsymbol{X}}} \right\|_\infty \leq C\{q_n(z) + a_n(z)\}.$$

In particular,

$$\|\hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}}\|_\infty \leq \|\boldsymbol{\Sigma}_{\tilde{\boldsymbol{X}}}\|_\infty + C\{q_n(z) + a_n(z)\} \leq C.$$

It follows that

$$\begin{aligned} \left\| \mathbf{e}_j^\top - \hat{\boldsymbol{\Theta}}_j^\top \hat{\boldsymbol{\Sigma}}_{\check{\boldsymbol{X}}} \right\|_\infty &\leq C r_{\Theta,j}(z) + C\sqrt{\tilde{s}_j}\{q_n(z) + a_n(z)\} \\ &\leq C \left\{ r_{\Theta,j}(z) + \tilde{s}_j q_n(z) + \tilde{s}_j^{3/2} a_n(z) \right\} \\ &= C u_{\Theta,j}(z), \end{aligned} \tag{55}$$

where the second inequality uses $\tilde{s}_j \geq 1$.

All the preceding bounds hold on the intersection of a fixed number of events, which, by a union bound, has probability at least $1 - Ce^{-cz}$. We now bound $E_0, \ldots, E_6$ on this event.

For E_0, Holder's inequality, (51), and (53) give

$$|E_0| \leq \|\hat{\boldsymbol{\Theta}}_j - \boldsymbol{\Theta}_j\|_1 \left\| \frac{1}{n} \tilde{\boldsymbol{X}}^\top \boldsymbol{\varepsilon} \right\|_\infty \leq C r_{\Theta,j}(z) q_n(z).$$

For E_1, Holder's inequality, (52), and (47) give

$$|E_1| \leq \|\hat{\boldsymbol{\Theta}}_j\|_1 \left\| \frac{1}{n} (\tilde{\boldsymbol{X}} - \check{\boldsymbol{X}})^\top \boldsymbol{\varepsilon} \right\|_\infty \leq C \tilde{s}_j a_n(z).$$

For E_2, using (45) and (55),

$$|E_2| \leq \left\| \mathbf{e}_j - \frac{1}{n} \hat{\boldsymbol{\Theta}}_j^\top \check{\boldsymbol{X}}^\top \check{\boldsymbol{X}} \right\|_\infty \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \leq C s_0 q_n(z) u_{\Theta,j}(z).$$

For E_3, by Holder's inequality, (45), (52), and (48),

$$|E_3| \leq \|\hat{\boldsymbol{\Theta}}_j\|_1 \left\| \frac{1}{n} \check{\boldsymbol{X}}^\top (\check{\boldsymbol{X}} - \tilde{\boldsymbol{X}}) \right\|_\infty \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \leq C s_0 \tilde{s}_j a_n(z) q_n(z).$$

For E_4, using (45), (52), and (54),

$$|E_4| \le \|\hat{\boldsymbol{\Theta}}_j\|_1 \left\| \frac{1}{n} \tilde{\boldsymbol{X}}^\top \boldsymbol{f}(\boldsymbol{W}) \right\|_\infty \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \le C s_0 \tilde{s}_j q_n(z)^2.$$

For E_5, since $\check{\boldsymbol{X}} - \tilde{\boldsymbol{X}} = \boldsymbol{f}(\boldsymbol{W}) - \hat{\boldsymbol{f}}(\boldsymbol{W})$, Cauchy's inequality gives

$$|E_5| \le \|\hat{\boldsymbol{\Theta}}_j\|_1 \max_{1 \le \ell \le p} \|\hat{\boldsymbol{f}}_\ell(\boldsymbol{W}) - \boldsymbol{f}_\ell(\boldsymbol{W})\|_n \|\boldsymbol{f}(\boldsymbol{W})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_n.$$

Using (43), (46), and (52), we obtain

$$|E_5| \le C \tilde{s}_j a_n(z)^{1/2} \sqrt{s_0} q_n(z).$$

Since $a_n(z) \ge q_n(z)^2$, we have

$$|E_5| \le C \sqrt{s_0}\, \tilde{s}_j a_n(z),$$

For E_6, using (52) and (50),

$$|E_6| \le \|\hat{\boldsymbol{\Theta}}_j\|_1 \left\| \frac{1}{n} \check{\boldsymbol{X}}^\top \{\hat{\boldsymbol{g}}(\boldsymbol{W}) - \boldsymbol{g}(\boldsymbol{W})\} \right\|_\infty \le C s_0 \tilde{s}_j a_n(z).$$

□

Now we establish the exponential integrability of the null statistic. Recall that $\mathcal{J}_0 = \{j \in [p] : \beta_j^{(0)} = 0\}$. For $j \in \mathcal{J}_0$, define

$$T_{n,j} := \sqrt{n_0} \hat{\beta}_j^{(dp)} = \sqrt{n_0} \left(\hat{\beta}_j^{(dp)} - \beta_j^{(0)} \right).$$

The linear expansion established in the proof of Theorem 2 gives

$$T_{n,j} = S_{n,j} + R_{n,j} + G_{n,j}, \qquad S_{n,j} := \frac{1}{\sqrt{n_0}} \sum_{i=1}^{n_0} \psi_{ij}, \qquad G_{n,j} := \sqrt{n_0} \tilde{\Delta}_j,$$

where

$$\psi_{ij} := \boldsymbol{\Theta}_j^\top \tilde{\boldsymbol{X}}_i^{(0)} \varepsilon_i^{(0)}, \qquad \tilde{\boldsymbol{X}}_i^{(0)} := \boldsymbol{X}_i^{(0)} - \mathbb{E}\{\boldsymbol{X}_i^{(0)} \mid \boldsymbol{W}_i^{(0)}\}.$$

Here $R_{n,j}$ is the debiasing remainder and $G_{n,j}$ is the contribution of the Gaussian privacy noise. We construct

$$\hat{e}_j = \frac{1}{2} \left[\exp\left\{ T_{n,j} - \frac{(\hat{V}_j^{(dp)})^2}{2} \right\} + \exp\left\{ -T_{n,j} - \frac{(\hat{V}_j^{(dp)})^2}{2} \right\} \right].$$

Lemma 5. Suppose that the conditions of Theorem 2 hold. If $\hat{V}_j^{(dp)} \xrightarrow{P} V_j$, where $V_j^2 = (\mathbf{\Theta})_{j,j}\sigma^2$, then

$$\mathbb{E}(\hat{e}_j) \to 1, \qquad j \in \mathcal{J}_0.$$

Consequently, $\hat{e}_j$ is an asymptotic e-value for every $j \in \mathcal{J}_0$.

Proof. We show this by proving the terms $\exp\left\{T_{n,j} - (\hat{V}_j^{(dp)})^2/2\right\}$ are uniformly integrable. Since $0 \leq \exp\{-(\hat{V}_j^{(dp)})^2/2\} \leq 1$, the products $\exp(\pm T_{n,j})\exp\{-(\hat{V}_j^{(dp)})^2/2\}$ are uniformly integrable whenever $\exp(\pm T_{n,j})$ are uniformly integrable. Based on the output of $\hat{\boldsymbol{\beta}}$ of Algorithm 1, condition on the event $\mathcal{E}_1' \cap \mathcal{E}_2'$, the untruncated expansion gives

$$R_{n,j} = \sqrt{n}\,(E_0 - E_1 + E_2 + E_3 - E_4 - E_5 - E_6).$$

Applying Lemma 4 with the probability parameter $z = t/(C\rho_n) - 1 \geq 1$, define

$$\rho_{n,j}(z) = \sqrt{n}\big[r_{\Theta,j}(z)q_n(z) + \tilde{s}_j a_n(z) + s_0 q_n(z) u_{\Theta,j}(z) + s_0\tilde{s}_j a_n(z) q_n(z) + s_0\tilde{s}_j q_n(z)^2 + \sqrt{s_0}\,\tilde{s}_j a_n(z) + s_0\tilde{s}_j a_n(z)\big].$$

The rate conditions in Theorem 2 imply

$$\rho_n := \sup_{j\in\mathcal{J}_0} \rho_{n,j}(1) = o(1).$$

Then the exponential envelope implies

$$\sup_{j\in\mathcal{J}_0} \mathbb{P}\left(|R_{n,j}| > t\right) \leq C\exp\left(-\frac{ct}{\rho_n}\right), \qquad t \geq 2C\rho_n,$$

after adjusting the constants. Splitting the tail integral at $2C\rho_n$, and using $\rho_n = o(1)$, we obtain, for every fixed $a > 0$,

$$\sup_{j\in\mathcal{J}_0} \mathbb{E}\exp\left\{a|R_{n,j}|\right\} \leq 1 + C\rho_n = 1 + o(1).$$

By Assumptions 1 and 4, there exist constants $C, c > 0$ such that

$$\sup_{j\in[p]} \mathbb{E}\exp(\varpi\psi_{ij}) \leq \exp(C\varpi^2), \qquad |\varpi| \leq c.$$

Hence, for every fixed $a > 0$, $a/\sqrt{n_0} \leq c$ for all sufficiently large n_0, the independence yields

$$\sup_{j\in\mathcal{J}_0} \mathbb{E}\exp(\pm a S_{n,j}) \leq \exp(Ca^2).$$

It follows that

$$\limsup_{n_0,p\to\infty} \sup_{j\in\mathcal{J}_0} \mathbb{E}\exp\{a|S_{n,j}|\} \leq 2\exp(Ca^2) < \infty.$$

Since $G_{n,j} \sim \mathrm{N}(0, \tau_n^2)$ with $\tau_n^2 = O(1)$ under the privacy-rate condition in Theorem 2, the Gaussian moment-generating function and $e^{a|x|} \leq e^{ax} + e^{-ax}$ imply that, for every fixed $a > 0$,

$$\limsup_{n_0, p \to \infty} \sup_{j \in \mathcal{J}_0} \mathbb{E} \exp\{a|G_{n,j}|\} < \infty.$$

For some $\kappa > 0$, and take $\alpha_S, \alpha_R, \alpha_G > 1$ such that

$$\frac{1}{\alpha_S} + \frac{1}{\alpha_R} + \frac{1}{\alpha_G} = 1,$$

Hölder's inequality gives

$$\mathbb{E} \exp\{(1+\kappa)|T_{n,j}|\} \leq [\mathbb{E} \exp\{\alpha_S(1+\kappa)|S_{n,j}|\}]^{1/\alpha_S} [\mathbb{E} \exp\{\alpha_R(1+\kappa)|R_{n,j}|\}]^{1/\alpha_R} \times [\mathbb{E} \exp\{\alpha_G(1+\kappa)|G_{n,j}|\}]^{1/\alpha_G}.$$

Therefore,

$$\limsup_{n_0, p \to \infty} \sup_{j \in \mathcal{J}_0} \mathbb{E} \exp\{(1+\kappa)|T_{n,j}|\} < \infty.$$

By the de la Vallée-Poussin criterion (Hu and Rosalsky, 2011), $\{\exp(T_{n,j}) : j \in \mathcal{J}_0\}$ and $\{\exp(-T_{n,j}) : j \in \mathcal{J}_0\}$ are uniformly integrable.

Theorem 2 gives $T_{n,j} \xrightarrow{D} N(0, V_j^2)$. Uniform integrability implies $\mathbb{E}\{\exp(T_{n,j})\} \to \exp(V_j^2/2)$, $\mathbb{E}\{\exp(-T_{n,j})\} \to \exp(V_j^2/2)$. Since $\hat{V}_j^{(dp)} \xrightarrow{P} V_j$ and $\exp\{-(\hat{V}_j^{(dp)})^2/2\} \leq 1$, Slutsky's theorem and the same uniform-integrability argument yield $\mathbb{E} \exp\left\{T_{n,j} - (\hat{V}_j^{(dp)})^2/2\right\} \to 1$, $\mathbb{E} \exp\left\{-T_{n,j} - (\hat{V}_j^{(dp)})^2/2\right\} \to 1$. Therefore $\mathbb{E}(\hat{e}_j) \to 1$, and $\hat{e}_j$ is an asymptotic e-value for every $j \in \mathcal{J}_0$. □

A4.7. Proof of Theorem 3

Proof. For the privacy guarantee, the candidate set is constructed as $\widehat{\mathcal{S}} = \mathrm{supp}(\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}$ is the preliminary private sparse estimator from Algorithm 1; hence $|\widehat{\mathcal{S}}| \leq s'$. Algorithm 3 releases private statistics only for variables in $\widehat{\mathcal{S}}$. For each selected variable $j \in \widehat{\mathcal{S}}$, the coefficient and variance estimators are released as one joint private pair $(\hat{\beta}_j^{(dp)}, \hat{V}_j^{(dp)})$ with privacy parameters $(\epsilon/(s'+1), \delta/(s'+1))$. The e-value $\hat{e}_j$ is then a deterministic post-processing of this pair, and $\hat{e}_j = 0$ is assigned for $j \notin \widehat{\mathcal{S}}$. Therefore, by sequential composition over one candidate-set release and at most s' joint pair releases, followed by post-processing for computing the e-values and applying the e-BH threshold, Algorithm 3 satisfies (ϵ, δ)-STDP. By Lemma 5, for each $j \in \mathcal{J}_0$, $\hat{e}_j$ is an asymptotic e-value, i.e., $\limsup_{n_0, p \to \infty} \mathbb{E}(\hat{e}_j) \leq 1$. Let t_q be the threshold selected by the e-BH rule in Algorithm 3, and let

$$R(t_q) = \sum_{j=1}^{p} \mathbb{I}\{\hat{e}_j \geq t_q\}$$

be the corresponding number of rejections. When $R(t_q) > 0$, the threshold definition gives $t_q R(t_q)/p \geq 1/q$; when $R(t_q) = 0$, the numerator in the false discovery proportion is zero. Hence,

$$\mathrm{FDR} = \mathbb{E}\left(\frac{\sum_{j\in\mathcal{J}_0} \mathbb{I}\{\hat{e}_j \geq t_q\}}{R(t_q) \vee 1}\right) \leq \frac{q}{p}\mathbb{E}\left(\sum_{j\in\mathcal{J}_0} \hat{e}_j\right).$$

Therefore,

$$\limsup_{n_0,p\to\infty} \mathrm{FDR} \leq \limsup_{n_0,p\to\infty} \frac{q}{p}\mathbb{E}(\sum_{j\in\mathcal{J}_0} \hat{e}_j) \leq q.$$

□

References

Cai, Z., S. Li, X. Xia, and L. Zhang (2026). Differentially private estimation and inference in high-dimensional regression with FDR control. *Journal of Machine Learning Research 27*, 1–54.

Hu, T.-C. and A. Rosalsky (2011). A note on the de La Vallée Poussin criterion for uniform integrability. *Statistics & Probability Letters 81*(1), 169–174.

Li, M., Y. Tian, Y. Feng, and Y. Yu (2024). Federated transfer learning with differential privacy. *arXiv preprint arXiv:2403.11343*.

Zhu, Y. (2017). Nonasymptotic analysis of semiparametric regression models with high-dimensional parametric coefficients. *The Annals of Statistics 45*(5), 2274–2298.

Zhu, Y., Z. Yu, and G. Cheng (2019). High dimensional inference in partially linear models. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2760–2769. PMLR.