A. Proofs

A.1. Proof of Lemma 1

Proof. In the following, we abbreviate $j$ in Lemma 1 for the simplicity of the notation unless there is no confusion, and prove the lemma in slightly general case of $V[y] = \Sigma$.

To prove the lemma, we first state the polyhedral lemma in Lee et al. (2016) as follows:

**Lemma 7** (Polyhedral Lemma; (Lee et al., 2016)). Suppose $y \sim N(\mu, \Sigma)$. Let $c = \Sigma \eta (\Sigma \eta)^{-1}$ for any $\eta \in \mathbb{R}^n$, and let $z = (I_n - cn^\top) y$. Then we have

$$
\text{Pol}(S) = \{ y \in \mathbb{R}^n \mid A y \leq b \}
= \left\{ y \in \mathbb{R}^n \mid L(S, z) \leq \eta^\top y \leq U(S, z), \right\},
$$

where

$$
L(S, z) = \max_{j : (A c)_{j} < 0} \left( b_j - (A z)_{j} \right), \quad \text{(13a)}
$$

$$
U(S, z) = \min_{j : (A c)_{j} > 0} \left( b_j - (A z)_{j} \right). \quad \text{(13b)}
$$

and $N(S, z) = \max_{j : (A c)_{j} = 0} b_j - (A z)_{j}$. In addition, $(L(S, z), U(S, z), N(S, z))$ is independent of $\eta^\top y$.

The polyhedral lemma allows us to construct a pivotal quantity as a truncated normal distribution, that is, for any $z$, we have

$$
[F_{\eta^\top y, \Sigma \eta}^{L(S, z), U(S, z)}(\eta^\top y) \mid y \in \text{Pol}(S)] \sim \text{Unif}(0, 1), \quad (14)
$$

where Unif$(0, 1)$ denotes the standard (continuous) uniform distribution. In fact, by letting $z_0$ be an arbitrary realization of $z$, one can see that

$$
[\eta^\top y \mid y \in \text{Pol}(S), z = z_0] = \frac{[\eta^\top y \mid L(S, z_0) \leq \eta^\top y \leq U(S, z_0)]}{\text{TN}(\eta^\top \mu, \eta^\top \Sigma \eta, L(S, z_0), U(S, z_0))},
$$

where $\frac{\cdot}{\cdot}$ denotes the equality of random variables in distribution. Therefore, probability integral transformation implies

$$
[F_{\eta^\top y, \Sigma \eta}^{L(S, z), U(S, z)}(\eta^\top y) \mid y \in \text{Pol}(S), z = z_0]
$$

has a uniform distribution Unif$(0, 1)$ for any $z_0$. By integrating out $z_0$, the pivotal quantity Eq.(14) holds. In addition, an lower $\alpha$-percentile of the distribution can be obtained as

$$
q_\alpha = (F_{\eta^\top y, \Sigma \eta}^{L(S, z), U(S, z)}(\eta^\top y) \mid y \in \text{Pol}(S), z = z_0)^{-1}(\alpha).
$$

In the following, let us denote $(S, z)$ by $S$ for shorthand. The remaining is to show that truncation points in Eqs.(13) are equivalent to

$$
L(S) = \eta^\top y + \theta_L \eta^\top \Sigma \eta \quad \text{(15a)}
$$

where $\theta_L = \min_{\theta \in \mathbb{R}} \theta$ s.t. $y + \theta \Sigma \eta \in \text{Pol}(S)$

and

$$
U(S) = \eta^\top y + \theta_U \eta^\top \Sigma \eta \quad \text{(15b)}
$$

where $\theta_U = \max_{\theta \in \mathbb{R}} \theta$ s.t. $y + \theta \Sigma \eta \in \text{Pol}(S),$

respectively. Simple calculation shows that, for any $\theta \in \mathbb{R}$, we have

$$
y + \theta \Sigma \eta \in \text{Pol}(S) \quad \iff \quad A(y + \theta \Sigma \eta) \leq b \quad \iff \quad \theta \cdot A \Sigma \eta \leq b - A y.
$$

On the other hand, by the definition of $c$ and $z$ in Lemma 7, it is easy to see that

$$
L(S) = \eta^\top y + \eta^\top \Sigma \eta \max_{j : (A \Sigma \eta)_j < 0} (b - A y)_j
$$

Therefore, for each $j$ such that $(A \Sigma \eta)_j < 0$, we have

$$
\max_{j : (A \Sigma \eta)_j < 0} \frac{(b - A y)_j}{(A \Sigma \eta)_j} \leq \theta
$$

and thus the minimum possible feasible $\theta$ would be

$$
\theta_L = \min \{ \theta \in \mathbb{R} \mid y + \theta \Sigma \eta \in \text{Pol}(S) \}
= \max_{j : (A \Sigma \eta)_j < 0} \frac{(b - A y)_j}{(A \Sigma \eta)_j}.
$$

Similarly, we see that the equivalency of $U(S)$.

To complete the proof, let us consider a Gaussian random variable $y$ with mean $X \beta^*$ and covariance matrix $\sigma^2 I_n$ with some constant $\sigma^2$. We can choose $\eta = (X_S^\top)^{-1} e_j$ for testing the null hypothesis $H_{0,j} : \beta_{S,j}^* = 0$ for each $j \in S$. Since $\eta^\top y$ reduces to the $j$-th element of an ordinary least square estimator for the selected model, and in this case, $\eta^\top \Sigma \eta$ reduces to

$$
\sigma^2_S = \sigma^2 \|\eta\|^2 = \sigma^2 (X^\top S X_S)_{jj}^{-1}.
$$

Then the critical values are computed as

$$
\ell_{\alpha/2} = q_{\alpha/2} = (F_{\eta^\top y, \Sigma \eta}^{L(S), U(S)}(\eta^\top y) \mid y \in \text{Pol}(S))^{-1}(\alpha/2).
$$
and
\[ u_{\alpha/2}^S = q_{1 - \alpha/2} = (F_{0, \sigma_3^2}^{[S(L)]})^{-1}(1 - \alpha/2), \]
respectively. From the above argument, there are no matter to compute the truncation points in Eqs.(15) based on the observations. In this case, Eqs.(15) can be written as
\[ L(S) = \eta^\top y + \theta_L \sigma^2(X_S^\top X_S)^{-1} \]
where \( \theta_L = \min_{\theta \in \mathbb{R}} \) s.t. \( y + \sigma^2(X_S^\top e_j \in \text{Pol}(S) \)
and
\[ U(S) = \eta^\top y + \theta_U \sigma^2(X_S^\top X_S)^{-1} \]
where \( \theta_U = \max_{\theta \in \mathbb{R}} \) s.t. \( y + \sigma^2(X_S^\top e_j \in \text{Pol}(S) \)
respectively, but we can ignore the scaling factor \( \sigma^2 \) because
\[ \min\{ \theta \in \mathbb{R}^n \mid y + \theta(X_S^\top e_j \in \text{Pol}(S) \}\]
\[ = \min\{ \sigma^2 \theta \in \mathbb{R}^n \mid y + \sigma^2(X_S^\top e_j \in \text{Pol}(S) \} \]
and
\[ \max\{ \theta \in \mathbb{R}^n \mid y + \theta(X_S^\top e_j \in \text{Pol}(S) \}
\[ = \max\{ \sigma^2 \theta \in \mathbb{R}^n \mid y + \sigma^2(X_S^\top e_j \in \text{Pol}(S) \} \].

A.2. Proof of Lemma 3

Proof. Since \( x_{ij} \in [0, 1] \), for any pair \((j, \tilde{j})\) such that \( \tilde{j} \in \text{Des}(j) \), \( x_{ij} \geq x_{\tilde{j}j} \) holds. Then,
\[
|x_{ij}^\top y| = | \sum_{i: y_i > 0} x_{ij} y_i + \sum_{i: y_i < 0} x_{ij} y_i |
\leq \max \left\{ \sum_{i: y_i > 0} x_{ij} y_i, - \sum_{i: y_i < 0} x_{ij} y_i \right\}
\leq \max \left\{ \sum_{i: y_i > 0} x_{ij} y_i, - \sum_{i: y_i < 0} x_{ij} y_i \right\}.
\]

A.3. Proof of Lemma 4

Proof. In MS, from Eq.(9), the constraint \( y + \theta \eta \in \text{Pol}(S) \)
is written as
\[
(-s_j x_{ij} - x_{j'j'})^\top (y + \theta \eta) \leq 0
\]
\[
\Leftrightarrow \frac{-(s_j x_{ij} + x_{j'j'})^\top y}{(s_j x_{ij} + x_{j'j'})^\top \eta} \leq \theta \text{ if } (s_j x_{ij} + x_{j'j'})^\top \eta > 0,
\]
and
\[
-\frac{(s_j x_{ij} + x_{j'j'})^\top y}{(s_j x_{ij} + x_{j'j'})^\top \eta} \geq \theta \text{ if } (s_j x_{ij} + x_{j'j'})^\top \eta < 0.
\]

Similarly, the conditions in Eqs.(16b), (16d), and (16f) imply that \( \theta_L = -\min\{ \theta^{(a)}_L, \theta^{(b)}_L, \theta^{(c)}_L \} \).

A.4. Proof of Lemma 5

Proof. First, note that \( 0 \leq x_{ij} \leq x_{ij'} \leq 1 \) for any \((j, j', j'') \in S \times S \times \text{Des}(j') \). We first prove Eq.(12a).
\[
(s_j x_{ij} + x_{j'j'})^\top y = s_j x_{ij}^\top y + \sum_{i: y_i > 0} x_{ij} y_i + \sum_{i: y_i < 0} x_{ij} y_i
\geq s_j x_{ij}^\top y + \sum_{i: y_i < 0} x_{ij} y_i
\geq s_j x_{ij}^\top y + \sum_{i: y_i < 0} x_{ij} y_i = L^{(a)}_E.
\]
which proves the first line. Next, we prove Eq. (12b).

\[
(s_j \mathbf{x}_j + \mathbf{x}_{j'})^T \mathbf{y} = s_j \mathbf{x}_j^T \mathbf{y} + \sum_{i:y_i > 0} x_{ij}y_i + \sum_{i:y_i < 0} x_{ij}y_i
\leq s_j \mathbf{x}_j^T \mathbf{y} + \sum_{i:y_i > 0} x_{ij}y_i.
\leq s_j \mathbf{x}_j^T \mathbf{y} + \sum_{i:y_i > 0} x_{ij}y_i = U_E^{(a)},
\]

which proves the second line. Eqs. (12c) to (12h) are proved similarly.

A.5. Proof of Theorem 6

Proof. First, we prove (i). For any \((j, j', j'') \in S \times S \times Des_j(j'')\), by using Lemma 5 directly, a lower and an upper bound of \(s_j \mathbf{x}_j^T \mathbf{y} + \mathbf{x}_{j'}^T \mathbf{y}\) can be obtained as

\[
L_E^{(a)} \leq s_j \mathbf{x}_j^T \mathbf{y} + \mathbf{x}_{j'}^T \mathbf{y} \leq U_E^{(a)}
\]

Similarly, a lower and an upper bound of \(s_j \mathbf{x}_j^T \eta + \mathbf{x}_{j'}^T \eta\) can be also obtained as

\[
L_D^{(a)} \leq s_j \mathbf{x}_j^T \eta + \mathbf{x}_{j'}^T \eta \leq U_D^{(a)}
\]

From Eq. (18), we have

\[
U_D^{(a)} < 0 \Rightarrow (s_j \mathbf{x}_j + \mathbf{x}_{j'})^T \eta < 0
\]

for all \((j, j') \in S \times Des_j(j')\). It means that the \((j, j')\)-th constraint does not affect the solution of the optimization problem in Eq. (11a). Now, we consider the case of \(U_D^{(a)} > 0\). If \(L_D^{(a)} > 0\), the value

\[
(s_j \mathbf{x}_j + \mathbf{x}_{j'})^T \mathbf{y} / (s_j \mathbf{x}_j + \mathbf{x}_{j'})^T \eta
\]

can be bounded below by \(L_E^{(a)} / U_E^{(a)}\) when \(L_E^{(a)} > 0\), and \(L_E^{(a)} / L_D^{(a)}\) when \(L_E^{(a)} < 0\), while the value can take any small values if \(L_D^{(a)} < 0\). As a result, for the current optimal solution \(\hat{\theta}_L^{(a)}\), \((j, j')\)-th constraint does not affect the solution of the optimization problem Eq. (11a), if

\[
L_D^{(a)} > 0, L_E^{(a)} > 0 \quad \text{and} \quad \frac{L_E^{(a)}}{U_D^{(a)}} > \hat{\theta}_L^{(a)},
\]

or

\[
L_D^{(a)} > 0, L_E^{(a)} < 0 \quad \text{and} \quad \frac{L_E^{(a)}}{L_D^{(a)}} > \hat{\theta}_L^{(a)},
\]

because \(L_D^{(a)} > 0\) implies \(U_D^{(a)} > 0\). Similarly, we can prove (ii) – (iv) by the same argument.

B. Selectivity inference for OMP

Lemma 8. Let \(\eta := (X^+)^T \mathbf{e}_j\). The solutions of the optimization problems in (7) are respectively written as

\[
\theta_L^{(a)} := \min_{h \in [k], j' \in S_h} \frac{(s_{(k)}X_{(k)} + x_{j'})^T P_{S_h} \mathbf{y}}{(s_{(k)}X_{(k)} + x_{j'})^T P_{S_h} \eta},
\]

\[
\theta_U^{(a)} := \max_{h \in [k], j' \in S_h} \frac{(s_{(k)}X_{(k)} - x_{j'})^T P_{S_h} \mathbf{y}}{(s_{(k)}X_{(k)} - x_{j'})^T P_{S_h} \eta}
\]

where

\[
\theta_L = -\min\{\theta_L^{(a)}, \theta_U^{(a)}, \theta_L^{(c)}\},
\]

\[
\theta_U = -\max\{\theta_U^{(a)}, \theta_U^{(b)}, \theta_U^{(c)}\},
\]

Selective Inference for Sparse High-Order Interaction Models
Lemma 9. For any \( h \in [k] \) and \((j', \tilde{j}') \in \hat{S}_h \times Des_{\hat{h}}(j')\),
\[
\begin{align*}
L^{(a)}_{E} & := s(h)x^{\top}_{(h)}P_{Sh}y + \sum_{i:|P_{Sh}y| > 0} x_{ij'}[P_{Sh}y]_i \\
& \leq (s(h)x_+(h) + x_{\tilde{j}'})^{\top}P_{Sh}y, \\
U^{(a)}_{E} & := s(h)x^{\top}_{(h)}P_{Sh}y + \sum_{i:|P_{Sh}y| > 0} x_{ij'}[P_{Sh}y]_i \\
& \geq (s(h)x_+(h) + x_{\tilde{j}'})^{\top}P_{Sh}y, \\
L^{(a)}_{D} & := s(h)x^{\top}_{(h)}\eta + \sum_{i:|P_{Sh}\eta| > 0} x_{ij'}[P_{Sh}\eta]_i \\
& \leq (s(h)x_+(h) + x_{\tilde{j}'})^{\top}P_{Sh}\eta, \\
U^{(a)}_{D} & := s(h)x^{\top}_{(h)}\eta + \sum_{i:|P_{Sh}\eta| > 0} x_{ij'}[P_{Sh}\eta]_i \\
& \geq (s(h)x_+(h) + x_{\tilde{j}'})^{\top}P_{Sh}\eta.
\end{align*}
\]

(iii) Furthermore, consider solving the optimization problem in Eq. (19d), and let \( \hat{\theta}^{(a)}_U \) be the current optimal solution. If
\[
\{L^{(a)}_{D} > 0\} \cup \{U^{(a)}_{D} < 0, L^{(a)}_{E} < 0, L^{(a)}_{E}/U^{(a)}_{D} > \hat{\theta}^{(a)}_U\}
\]
\[
\cup \{U^{(a)}_{D} < 0, L^{(a)}_{E} > 0, L^{(a)}_{E}/U^{(a)}_{D} > \hat{\theta}^{(a)}_U\}
\]
is true, then the \( \tilde{j}'\)-th constraint in Eq. (10a) for any \( h \in [k] \) and \((j', \tilde{j}') \in \hat{S}_h \times Des_{\hat{h}}(j')\) does not affect the optimal solution in Eq. (19d).

(iv) Finally, consider solving the optimization problem in Eq. (19e), and let \( \hat{\theta}^{(b)}_U \) be the current optimal solution. If
\[
\{L^{(b)}_{D} > 0\} \cup \{U^{(b)}_{D} < 0, L^{(b)}_{E} < 0, L^{(b)}_{E}/U^{(b)}_{D} > \hat{\theta}^{(b)}_U\}
\]
\[
\cup \{U^{(b)}_{D} < 0, L^{(b)}_{E} > 0, L^{(b)}_{E}/U^{(b)}_{D} > \hat{\theta}^{(b)}_U\}
\]
is true, then the \( j'\)-th constraint in Eq. (10b) for any \( h \in [k] \) and \((j', \tilde{j}') \in \hat{S}_h \times Des_{\hat{h}}(j')\) does not affect the optimal solution in Eq. (19e).

Theorem 10. (i) Consider solving the optimization problem in Eq. (19a), and let \( \hat{\theta}^{(a)}_L \) be the current optimal solution, i.e., we know that the optimal \( \theta^{(a)}_L \) is at least no greater than \( \hat{\theta}^{(a)}_L \). If
\[
\{U^{(a)}_{D} < 0\} \cup \{L^{(a)}_{D} > 0, L^{(a)}_{E} < 0, L^{(a)}_{E}/L^{(a)}_{D} > \hat{\theta}^{(a)}_L\}
\]
\[
\cup \{L^{(a)}_{D} > 0, L^{(a)}_{E} > 0, L^{(a)}_{E}/L^{(a)}_{D} > \hat{\theta}^{(a)}_L\}
\]
is true, then the \( j'\)-th constraint in Eq. (10a) for any \( h \in [k] \) and \((j', \tilde{j}') \in \hat{S}_h \times Des_{\hat{h}}(j')\) does not affect the optimal solution in Eq. (19a).

(ii) Next, consider solving the optimization problem in Eq. (19b), and let \( \hat{\theta}^{(b)}_L \) be the current optimal solution. If
\[
\{U^{(b)}_{D} < 0\} \cup \{L^{(b)}_{D} > 0, L^{(b)}_{E} < 0, L^{(b)}_{E}/L^{(b)}_{D} > \hat{\theta}^{(b)}_L\}
\]
\[
\cup \{L^{(b)}_{D} > 0, L^{(b)}_{E} > 0, L^{(b)}_{E}/L^{(b)}_{D} > \hat{\theta}^{(b)}_L\}
\]
is true, then the \( \tilde{j}'\)-th constraint in Eq. (10b) for any \( h \in [k] \) and \((j', \tilde{j}') \in \hat{S}_h \times Des_{\hat{h}}(j')\) does not affect the optimal solution in Eq. (19b).