Supplemental Material: Safe Element Screening for Submodular Function Minimization

Weizhong Zhang[1] **Bin Hong**[2] **Lin Ma**[1] **Wei Liu**[1] **Tong Zhang**[1]

[1] Tencent AI Lab, Shenzhen, China [2]State Key Lab of CAD&CG, Zhejiang University, China

In this supplement, we present the detailed proofs of all the theorems in the main text.

A. Proof of Theorem 1

Proof. of Theorem 1:
(i) Since $f(\mathbf{w}) = \max_{\mathbf{s}\in B(F)}\langle \mathbf{w}, \mathbf{s}\rangle$, we can have

$$\begin{aligned}
&\min_{\mathbf{w}\in\mathbb{R}^p} f(\mathbf{w}) + \sum_{j=1}^{p}\psi_j([\mathbf{w}]_j) && (8)\\
=&\min_{\mathbf{w}\in\mathbb{R}^p}\max_{\mathbf{s}\in B(F)}\langle \mathbf{w}, \mathbf{s}\rangle + \sum_{j=1}^{p}\psi_j([\mathbf{w}]_j)\\
=&\max_{\mathbf{s}\in B(F)}\min_{\mathbf{w}\in\mathbb{R}^p}\langle \mathbf{w}, \mathbf{s}\rangle + \sum_{j=1}^{p}\psi_j([\mathbf{w}]_j) && (9)\\
=&\max_{\mathbf{s}\in B(F)} -\sum_{j=1}^{p}\psi_j^*(-[\mathbf{s}]_j), && (10)
\end{aligned}$$

where (9) holds due to the strong duality theorem (Borwein & Lewis, 2010), and (10) holds due to the definitions of the Fenchel conjugate of ψ_j.
(ii) From (8), we have

$$\begin{aligned}
&\mathbf{s}^* \in \arg\max_{\mathbf{s}\in B(F)}\langle \mathbf{w}^*, \mathbf{s}\rangle\\
&\Leftrightarrow \langle \mathbf{w}^*, \mathbf{s}^*\rangle \geq \langle \mathbf{w}^*, \mathbf{s}\rangle, \forall \mathbf{s}\in B(F)\\
&\Leftrightarrow \mathbf{w}^* \in N_{B(F)}(\mathbf{s}^*).
\end{aligned}$$

From Eq. (10), we have

$$\begin{aligned}
&\mathbf{w}^* \in \arg\min_{\mathbf{w}\in\mathbb{R}^p}\langle \mathbf{w}, \mathbf{s}^*\rangle + \sum_{j=1}^{p}\psi_j([\mathbf{w}]_j)\\
&\Leftrightarrow [\mathbf{s}]_k^* \in -\partial\psi_k([\mathbf{w}]_k^*), \forall k \in V.
\end{aligned}$$

The proof is complete. □

B. Proof of Lemma 1

Proof. of Lemma 1:
(i) It is the immediate conclusion of Theorem 2.
(ii) Since $\hat{\mathcal{E}} \subseteq A^*$ and $\hat{\mathcal{G}} \subseteq V/A^*$, we can solve the problem SFM by fixing the set $\hat{\mathcal{E}}$ and optimizing over $V/(\hat{\mathcal{E}}\cup\hat{\mathcal{G}})$. And the objective function becomes $\hat{F}(C) := F(\hat{\mathcal{E}}\cup C) - F(\hat{\mathcal{E}})$ with $C \subseteq V/(\hat{\mathcal{E}}\cup\hat{\mathcal{G}})$. Thus, SFM can be deduced to

$$\min_{C\subseteq V/(\hat{\mathcal{E}}\cup\hat{\mathcal{G}})} \hat{F}(C) := F(\hat{\mathcal{E}}\cup C) - F(\hat{\mathcal{E}}).$$

The second term of the new objective function $\hat{F}(C)$ is added to make $\hat{F}(\emptyset) = 0$, which is essential in submodular function analysis, such as Lovász extension, submodular and base polyhedra.
Below, we argue that $\hat{F}(C)$ is a submodular function.
For all $S \subseteq V/(\hat{\mathcal{E}} \cup \hat{\mathcal{G}})$ and $T \subseteq V/(\hat{\mathcal{E}} \cup \hat{\mathcal{G}})$, we have

$$\begin{aligned}
\hat{F}(S) + \hat{F}(T) &= (F(\hat{\mathcal{E}} \cup S) - F(\hat{\mathcal{E}})) + (F(\hat{\mathcal{E}} \cup T) - F(\hat{\mathcal{E}})) \\
&= F(\hat{\mathcal{E}} \cup S) + F(\hat{\mathcal{E}} \cup T) - 2F(\hat{\mathcal{E}}) \\
&\geq F((\hat{\mathcal{E}} \cup S) \cup (\hat{\mathcal{E}} \cup T)) + F((\hat{\mathcal{E}} \cup S) \cap (\hat{\mathcal{E}} \cup T)) - 2F(\hat{\mathcal{E}}) \qquad (11) \\
&= F(\hat{\mathcal{E}} \cup (S \cup T)) + F(\hat{\mathcal{E}} \cup (S \cup T)) - 2F(\hat{\mathcal{E}}) \\
&= (F(\hat{\mathcal{E}} \cup (S \cup T)) - F(\hat{\mathcal{E}})) + (F(\hat{\mathcal{E}} \cup (S \cup T)) - F(\hat{\mathcal{E}})) \\
&= \hat{F}(S \cup T) + \hat{F}(S \cap T).
\end{aligned}$$

The inequality (11) comes from the submoduality of F.
(iii) It is the immediate conclusion of (ii).
The proof is complete. □

C. Proof of Theorem 3

To prove Theorem 3, we need the following Lemma.

Lemma 4. [Dual of minimization of submodular functions, Proposition 10.3 in (Bach et al., 2013)] *Let F be a submodular function such that $F(\emptyset) = 0$. We have:*

$$\min_{A \subseteq V} F(A) = \max_{\mathbf{s} \in B(F)} \mathbf{s}_-(V) = \frac{1}{2}\Big(F(V) - \min_{\mathbf{s} \in B(F)} \|\mathbf{s}\|_1\Big), \qquad (12)$$

where $[\mathbf{s}_-]_k = \min\{[\mathbf{s}]_k, 0\}$ *for* $\forall k \in V$.

We now turn to prove Theorem 3.

Proof. of Theorem 3:
Since $\hat{P}(\hat{\mathbf{w}})$ is 1-strongly convex, for any $\hat{\mathbf{w}} \in dom\hat{P}(\hat{\mathbf{w}})$ and $\hat{\mathbf{w}}^* = \arg\min_{\hat{\mathbf{w}} \in \mathbb{R}^{\hat{p}}} \hat{P}(\hat{\mathbf{w}})$, we can have

$$\hat{P}(\hat{\mathbf{w}}) \geq \hat{P}(\mathbf{w}^*) + \langle \hat{\mathbf{g}}, \hat{\mathbf{w}} - \hat{\mathbf{w}}^* \rangle + \frac{1}{2}\|\hat{\mathbf{w}} - \hat{\mathbf{w}}^*\|_2^2,$$

where $\mathbf{g} \in \partial \hat{P}(\hat{\mathbf{w}}^*)$.
Since $dom\hat{P}(\hat{\mathbf{w}}) = \mathbb{R}^{\hat{p}}$, it holds that $0 \in \partial \hat{P}(\hat{\mathbf{w}}^*)$. Hence, we can obtain

$$\frac{1}{2}\|\hat{\mathbf{w}} - \hat{\mathbf{w}}^*\|_2^2 \leq \hat{P}(\hat{\mathbf{w}}) - \hat{P}(\hat{\mathbf{w}}^*).$$

In addition, we notice that $\hat{P}(\mathbf{w}^*) \geq \hat{D}(\hat{\mathbf{s}})$ for all $\hat{\mathbf{s}} \in B(\hat{F})$. By substituting this inequality into the above inequality, we obtain that

$$\frac{1}{2}\|\hat{\mathbf{w}} - \hat{\mathbf{w}}^*\|_2^2 \leq \hat{P}(\hat{\mathbf{w}}) - \hat{P}(\hat{\mathbf{w}}^*) \leq \hat{P}(\hat{\mathbf{w}}) - \hat{D}(\hat{\mathbf{s}}) = G(\hat{\mathbf{w}}, \hat{\mathbf{s}}).$$

Thus,

$$\hat{\mathbf{w}}^* \in \mathcal{B} := \Big\{\mathbf{w} : \|\mathbf{w} - \hat{\mathbf{w}}\| \leq \sqrt{2G(\hat{\mathbf{w}}, \hat{\mathbf{s}})}\Big\}. \qquad (13)$$

According to the equation (Opt) in Theorem 1, we have that $-\hat{\mathbf{w}}^*$ is the optimal solution of the problem Q-D'. Therefore, $-\hat{\mathbf{w}}^* \in B(\hat{F})$. From the definition of $B(\hat{F})$, we have

$$-\langle \hat{\mathbf{w}}^*, \mathbf{1} \rangle = -\hat{\mathbf{w}}^*(\hat{V}) = \hat{F}(\hat{V}).$$

Thus,

$$\hat{\mathbf{w}}^* \in \mathcal{P} := \Big\{ \mathbf{w} : \langle \mathbf{w}, \mathbf{1} \rangle = -\hat{F}(\hat{V}) \Big\}. \tag{14}$$

By section 7.3 of (Bach et al., 2013)), it holds that the unique minimizer of problem Q-D' is also a maximizer of

$$\max_{\mathbf{s} \in B(\hat{F})} \mathbf{s}_-(V).$$

Hence, it holds that

$$\|\hat{\mathbf{s}}^*\|_1 \leq \|\hat{\mathbf{s}}\|_1 \text{ for all } \hat{\mathbf{s}} \in B(\hat{F}). \tag{15}$$

From Lemma 4, we have

$$\begin{aligned} \hat{F}(C) &\geq \frac{1}{2}(\hat{F}(\hat{V}) - \|\hat{\mathbf{s}}\|_1), \text{for all } \hat{\mathbf{s}} \in B(\hat{F}), \\ \Rightarrow \|\hat{\mathbf{s}}\|_1 &\geq \hat{F}(\hat{V}) - 2\hat{F}(C), \text{for all } \hat{\mathbf{s}} \in B(\hat{F}). \end{aligned} \tag{16}$$

By combining (15) and (16), we acquire

$$\hat{F}(\hat{V}) - 2\hat{F}(C) \leq \|\hat{\mathbf{s}}^*\|_1 \leq \|\hat{\mathbf{s}}\|_1, \text{for all } \hat{\mathbf{s}} \in B(\hat{F}).$$

Since $\hat{\mathbf{w}}^* = -\hat{\mathbf{s}}^*$, we have

$$\hat{F}(\hat{V}) - 2\hat{F}(C) \leq \|\hat{\mathbf{w}}^*\|_1 \leq \|\hat{\mathbf{s}}\|_1, \text{for all } \hat{\mathbf{s}} \in B(\hat{F}).$$

Thus, we obtain

$$\hat{\mathbf{w}}^* \in \Omega := \Big\{ \mathbf{w} : \hat{F}(\hat{V}) - 2\hat{F}(C) \leq \|\mathbf{w}\|_1 \leq \|\hat{\mathbf{s}}\|_1 \Big\}. \tag{17}$$

From (13), (14) and (17), we have $\hat{\mathbf{w}}^* \in \mathcal{B} \cap \Omega \cap \mathcal{P}$.
The proof is complete. □

D. Proof of Lemma 2

Proof. of Lemma 2:
For any $j = 1, ..., \hat{p}$, we have

$$\sum_{i \neq j} ([\mathbf{w}]_i - [\hat{\mathbf{w}}]_i)^2 \leq 2G(\hat{\mathbf{w}}, \hat{\mathbf{s}}) - ([\mathbf{w}]_j - ([\hat{\mathbf{w}}]_j)^2, \tag{18}$$

$$\sum_{i \neq j} [\mathbf{w}]_i = -\hat{F}(\hat{V}) - [\mathbf{w}]_j. \tag{19}$$

By fixing the component $[\mathbf{w}]_j$, we can see that (18) and (19) are a ball and a plane in $\mathbb{R}^{\hat{p}-1}$, respectively. To make the intersection of (18) and (19) non-empty, we just need to restrict the distance between the center of ball (18) and the plane (19) smaller than the radius, *i.e.*,

$$\frac{|\sum_{i \neq j} [\hat{\mathbf{w}}]_i + \hat{F}(\hat{V}) + [\mathbf{w}]_j|}{\sqrt{\hat{p} - 1}} \leq \sqrt{2G(\hat{\mathbf{w}}, \hat{\mathbf{s}}) - ([\mathbf{w}]_j - [\hat{\mathbf{w}}]_j)^2},$$

which is equivalent to

$$\hat{p}[\mathbf{w}]_j^2 + b[\mathbf{w}]_j + c \leq 0, \tag{20}$$

$$\text{where } b = 2\Big(\sum_{i \neq j} [\hat{\mathbf{w}}]_i + \hat{F}(\hat{V}) - (\hat{p} - 1)[\hat{\mathbf{w}}]_j\Big), \text{ and } c = \Big(\sum_{i \neq j} [\hat{\mathbf{w}}]_i + \hat{F}(\hat{V})\Big)^2 - 2(\hat{p} - 1)G(\hat{\mathbf{w}}, \hat{\mathbf{s}}) + (\hat{p} - 1)[\hat{\mathbf{w}}]_j^2.$$

Thus we have

$$[\mathbf{w}]_j \in [\frac{-b-\sqrt{b^2-4\hat{p}c}}{2\hat{p}}, \frac{-b+\sqrt{b^2-4\hat{p}c}}{2\hat{p}}]$$

At last, we would point out here that since $\hat{\mathbf{w}}^*$ must be in the intersection of the ball (18) and the plane (19). Hence, inequality (20) can be satisfied with $[\hat{\mathbf{w}}^*]_j$, which implies that $b^2 - 4\hat{p}c$ would never be negative.
The proof is complete. □

E. Proof of Theorem 4

Proof. of Theorem 4:
(i): According to $\min_{\mathbf{w}\in\mathcal{B}\cap\mathcal{P}} [\mathbf{w}]_j = [\mathbf{w}]_j^{\min} > 0$ and $\hat{\mathbf{w}}^* \in \mathcal{B}\cap\mathcal{P}$, we have

$$[\hat{\mathbf{w}}^*]_j > 0.$$

From Theorem 2, it holds that $j \in \arg\min_{C\subseteq\hat{V}} \hat{F}(C) \subseteq A^*$.
(ii): Since $\max_{\mathbf{w}\in\mathcal{B}\cap\mathcal{P}} [\mathbf{w}]_j = [\mathbf{w}]_j^{\max} < 0$ and $\hat{\mathbf{w}}^* \in \mathcal{B}\cap\mathcal{P}$, we have

$$[\hat{\mathbf{w}}^*]_j < 0.$$

From Theorem 2, we have $j \notin \arg\min_{C\subseteq\hat{V}} \hat{F}(C)$. Note that $A^* = \mathcal{E} \cup \arg\min \hat{F}(C)$ and $j \notin \mathcal{E}$. Therefore $j \notin A^*$.
(iii) It is the immediate conclusion from (i) and (ii).
The proof is complete. □

F. Proof of Lemma 3

Proof. of Lemma 3:
(i) We just need to prove that

$$\begin{cases} [\mathbf{w}]_j^{\min} > 0 \text{ if } [\hat{\mathbf{w}}]_j > \sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}, \\ [\mathbf{w}]_j^{\max} < 0 \text{ if } [\hat{\mathbf{w}}]_j < -\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}. \end{cases}$$

We divide the proof into two parts. First, when $[\hat{\mathbf{w}}]_j > \sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}$, considering the definition of $[\mathbf{w}]_j^{\min}$, we have

$$[\mathbf{w}]_j^{\min} = \min_{\mathbf{w}\in\mathcal{B}\cap\mathcal{P}} [\mathbf{w}]_j \geq \min_{\mathbf{w}\in\mathcal{B}} [\mathbf{w}]_j = [\hat{\mathbf{w}}]_j - \sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})} > 0.$$

In this case, the element j can be screened by rule AES-1.
On the other hand, when $[\hat{\mathbf{w}}]_j < -\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}$, from the definition of $[\mathbf{w}]_j^{\max}$, we have

$$[\mathbf{w}]_j^{\max} = \max_{\mathbf{w}\in\mathcal{B}\cap\mathcal{P}} [\mathbf{w}]_j \leq \max_{\mathbf{w}\in\mathcal{B}} [\mathbf{w}]_j = [\hat{\mathbf{w}}]_j + \sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})} < 0.$$

In this case, the element j can be screened by rule IES-1.
(ii) We note that the point $\mathbf{v}$ with $[\mathbf{v}]_j = 0$ and $[\mathbf{v}]_k = [\hat{\mathbf{w}}]_k$ for all $k \neq j, k = 1, 2, .., \hat{p}$ belongs to the ball $\mathcal{B}$. Thus, we have

$$\min_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\leq 0} \|\mathbf{w}\|_1 \leq \sum_{i\neq j} |[\hat{\mathbf{v}}]_i| = \|\hat{\mathbf{w}}\|_1 - [\hat{\mathbf{w}}]_j < \|\hat{\mathbf{w}}\|_1.$$

Now, we turn to calculate $\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\leq 0} \|\mathbf{w}\|_1$.
We note that the range of $[\mathbf{w}]_j$ is $[-\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})} + [\hat{\mathbf{w}}]_j, \sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})} + [\hat{\mathbf{w}}]_j]$ when $\mathbf{w} \in \mathcal{B}$. Hence, the problem $\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\leq 0} \|\mathbf{w}\|_1$ can be decomposed into

$$\max_{-\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}+[\hat{\mathbf{w}}]_j\leq\alpha\leq 0} \left\{ \max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j=\alpha} \|\mathbf{w}\|_1 \right\}.$$

We assume $[\mathbf{w}]_j = \alpha$ with $-\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})} + [\hat{\mathbf{w}}]_j \leq \alpha \leq 0$ and first consider the following problem,

$$\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j=\alpha} \|\mathbf{w}\|_1,$$

which can be rewritten as

$$\begin{aligned} &\max_{[\mathbf{w}]_i, i\neq j} -\alpha + \sum_{i\neq j} |[\mathbf{w}]_i| \\ &s.t. \sum_{i\leq\hat{p}, i\neq j} ([\mathbf{w}]_i - [\hat{\mathbf{w}}]_j)^2 \leq 2G(\hat{\mathbf{w}},\hat{\mathbf{s}}) - (\alpha - [\hat{\mathbf{w}}]_j)^2. \end{aligned}$$

It is easy to check that the optimal solution of the problem above is

$$[\mathbf{w}]_i = [\hat{\mathbf{w}}]_i + \mathbf{sign}([\hat{\mathbf{w}}]_i)\sqrt{\frac{2G(\hat{\mathbf{w}},\hat{\mathbf{s}}) - (\alpha - [\hat{\mathbf{w}}]_j)^2}{\hat{p}-1}}.$$

The function $\mathbf{sign}(\cdot) : \mathbb{R} \to \{-1, 1\}$ above takes 1 if the argument is positive, otherwise takes -1. And the corresponding optimal value is

$$\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j=\alpha} \|\mathbf{w}\|_1 = -\alpha + \sum_{i\neq j} |[\hat{\mathbf{w}}]_i| + \sqrt{\hat{p}-1}\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}}) - (\alpha - [\hat{\mathbf{w}}]_j)^2}.$$

Now, we denote $h(\alpha) = -\alpha + \sum_{i\neq j} |[\hat{\mathbf{w}}]_i| + \sqrt{\hat{p}-1}\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}}) - (\alpha - [\hat{\mathbf{w}}]_j)^2}$ and turn to solve

$$\max_{-\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})} + [\hat{\mathbf{w}}]_j \leq \alpha \leq 0} h(\alpha)$$

If $[\hat{\mathbf{w}}]_j - \sqrt{\frac{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}{\hat{p}}} < 0$, then

$$\begin{aligned} \max_{-\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})} + [\hat{\mathbf{w}}]_j \leq \alpha \leq 0} h(\alpha) &= h([\hat{\mathbf{w}}]_j - \sqrt{\frac{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}{\hat{p}}}) = -[\hat{\mathbf{w}}]_j + \sum_{i\neq j} |[\hat{\mathbf{w}}]_i| + \sqrt{2\hat{p}G(\hat{\mathbf{w}},\hat{\mathbf{s}})} \\ &= \|\hat{\mathbf{w}}\|_1 - 2[\hat{\mathbf{w}}]_j + \sqrt{2\hat{p}G(\hat{\mathbf{w}},\hat{\mathbf{s}})}; \end{aligned}$$

else if $[\hat{\mathbf{w}}]_j - \sqrt{\frac{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}{\hat{p}}} \geq 0$, then

$$\begin{aligned} \max_{-\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})} + [\hat{\mathbf{w}}]_j \leq \alpha \leq 0} h(\alpha) = h(0) &= \sum_{i\neq j} |[\hat{\mathbf{w}}]_i| + \sqrt{\hat{p}-1}\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}}) - [\hat{\mathbf{w}}]_j^2} \\ &= \|\hat{\mathbf{w}}\|_1 - [\mathbf{w}]_j + \sqrt{\hat{p}-1}\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}}) - [\hat{\mathbf{w}}]_j^2}. \end{aligned}$$

In a consequence, we have

$$\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\leq 0} \|\mathbf{w}\|_1 = \begin{cases} \|\hat{\mathbf{w}}\|_1 - 2[\hat{\mathbf{w}}]_j + \sqrt{2\hat{p}G(\hat{\mathbf{w}},\hat{\mathbf{s}})}, & \text{if } [\hat{\mathbf{w}}]_j - \sqrt{\frac{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}{\hat{p}}} < 0 \\ \|\hat{\mathbf{w}}\|_1 - [\mathbf{w}]_j + \sqrt{\hat{p}-1}\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}}) - [\hat{\mathbf{w}}]_j^2}, & \text{otherwise.} \end{cases}$$

(iii) Recall that the point $\mathbf{v}$ with $[\mathbf{v}]_j = 0$ and $[\mathbf{v}]_k = [\hat{\mathbf{w}}]_k$ for all $k \neq j, k = 1, 2, .., \hat{p}$ lies in the ball $\mathcal{B}$. Thus, we have

$$\min_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\geq 0} \|\mathbf{w}\|_1 \leq \sum_{i\neq j} |[\hat{\mathbf{v}}]_i| = \|\hat{\mathbf{w}}\|_1 - [\hat{\mathbf{w}}]_j < \|\hat{\mathbf{w}}\|_1.$$

Now, we turn to calculate $\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\geq 0}\|\mathbf{w}\|_1$.

We note that the range of $[\mathbf{w}]_j$ is $[-\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}+[\hat{\mathbf{w}}]_j, \sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}+[\hat{\mathbf{w}}]_j]$ when $\mathbf{w}\in\mathcal{B}$. Hence, we decompose the problem $\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\geq 0}\|\mathbf{w}\|_1$ into

$$\max_{0\leq\alpha\leq\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}+[\hat{\mathbf{w}}]_j}\Big\{\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j=\alpha}\|\mathbf{w}\|_1\Big\}.$$

We assume $[\hat{\mathbf{w}}]_j=\alpha$ with $0\leq\alpha\leq\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}+[\hat{\mathbf{w}}]_j$ and first solve the following problem:

$$\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j=\alpha}\|\mathbf{w}\|_1,$$

which can be rewritten as

$$\begin{aligned}&\max_{[\mathbf{w}]_i,i\neq j}\ \alpha+\sum_{i\neq j}|[\mathbf{w}]_i|\\ &s.t.\sum_{i\leq\hat{p},i\neq j}([\mathbf{w}]_i-[\hat{\mathbf{w}}]_j)^2\leq 2G(\hat{\mathbf{w}},\hat{\mathbf{s}})-(\alpha-[\hat{\mathbf{w}}]_j)^2.\end{aligned}$$

It can be verified that the optimal solution of the problem above is

$$[\mathbf{w}]_i=[\hat{\mathbf{w}}]_i+\mathbf{sign}([\hat{\mathbf{w}}]_i)\sqrt{\frac{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})-(\alpha-[\hat{\mathbf{w}}]_j)^2}{\hat{p}-1}}.$$

The function $\mathbf{sign}(\cdot):\mathbb{R}\to\{-1,1\}$ above takes 1 if the argument is positive, otherwise takes -1. And the corresponding optimal value is

$$\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j=\alpha}\|\mathbf{w}\|_1=\alpha+\sum_{i\neq j}|[\hat{\mathbf{w}}]_i|+\sqrt{\hat{p}-1}\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})-(\alpha-[\hat{\mathbf{w}}]_j)^2}.$$

Now, we denote $h(\alpha)=\alpha+\sum_{i\neq j}|[\hat{\mathbf{w}}]_i|+\sqrt{\hat{p}-1}\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})-(\alpha-[\hat{\mathbf{w}}]_j)^2}$ and turn to solve

$$\max_{0\leq\alpha\leq\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}+[\hat{\mathbf{w}}]_j}h(\alpha)$$

If $[\hat{\mathbf{w}}]_j+\sqrt{\frac{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}{\hat{p}}}>0$, then

$$\begin{aligned}\max_{0\leq\alpha\leq\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}+[\hat{\mathbf{w}}]_j}h(\alpha)&=h([\hat{\mathbf{w}}]_j+\sqrt{\frac{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}{\hat{p}}})=[\hat{\mathbf{w}}]_j+\sum_{i\neq j}|[\hat{\mathbf{w}}]_i|+\sqrt{2\hat{p}G(\hat{\mathbf{w}},\hat{\mathbf{s}})}\\ &=\|\hat{\mathbf{w}}\|_1+2[\hat{\mathbf{w}}]_j+\sqrt{2\hat{p}G(\hat{\mathbf{w}},\hat{\mathbf{s}})}.\end{aligned}$$

Else if $[\hat{\mathbf{w}}]_j+\sqrt{\frac{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}{\hat{p}}}\leq 0$, then

$$\begin{aligned}\max_{0\leq\alpha\leq\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}+[\hat{\mathbf{w}}]_j}h(\alpha)&=h(0)=\sum_{i\neq j}|[\hat{\mathbf{w}}]_i|+\sqrt{\hat{p}-1}\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})-[\hat{\mathbf{w}}]_j^2}\\ &=\|\hat{\mathbf{w}}\|_1+[\mathbf{w}]_j+\sqrt{\hat{p}-1}\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})-[\hat{\mathbf{w}}]_j^2}.\end{aligned}$$

Consequently, we have

$$\max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\geq 0}\|\mathbf{w}\|_1=\begin{cases}\|\hat{\mathbf{w}}\|_1+2[\hat{\mathbf{w}}]_j+\sqrt{2\hat{p}G(\hat{\mathbf{w}},\hat{\mathbf{s}})}, & \text{if } [\hat{\mathbf{w}}]_j+\sqrt{\frac{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}{\hat{p}}}>0,\\ \|\hat{\mathbf{w}}\|_1+[\mathbf{w}]_j+\sqrt{\hat{p}-1}\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})-[\hat{\mathbf{w}}]_j^2}, & \text{otherwise.}\end{cases}$$

The proof is complete. □

G. Proof of Theorem 5

Proof. of Theorem 5:
(i): Noting that

$$\begin{cases}0<[\hat{\mathbf{w}}]_j\leq\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})},\\ \max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\leq 0}\|\mathbf{w}\|_1<\hat{F}(\hat{V})-2\hat{F}(C),\end{cases}$$

and $\Omega=\left\{\mathbf{w}:\hat{F}(\hat{V})-2\hat{F}(C)\leq\|\mathbf{w}\|_1\leq\|\hat{\mathbf{s}}\|_1\right\}$, we have

$$\left\{\mathbf{w},\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\leq 0\right\}\cap\Omega=\emptyset. \tag{21}$$

Since $\hat{\mathbf{w}}^*\in\mathcal{B}\cap\Omega$, from (21) we have $[\hat{\mathbf{w}}^*]_j>0$. Thus, from Theorem 2 we have $j\in\arg\min\hat{F}(C)\subseteq A^*$.
(ii): Since

$$\begin{cases}-\sqrt{2G(\hat{\mathbf{w}},\hat{\mathbf{s}})}\leq[\hat{\mathbf{w}}]_j<0,\\ \max_{\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\geq 0}\|\mathbf{w}\|_1<\hat{F}(\hat{V})-2\hat{F}(C),\end{cases}$$

and $\Omega=\left\{\mathbf{w}:\hat{F}(\hat{V})-2\hat{F}(C)\leq\|\mathbf{w}\|_1\leq\|\hat{\mathbf{s}}\|_1\right\}$, we have

$$\left\{\mathbf{w},\mathbf{w}\in\mathcal{B},[\mathbf{w}]_j\geq 0\right\}\cap\Omega=\emptyset. \tag{22}$$

Since $\hat{\mathbf{w}}^*\in\mathcal{B}\cap\Omega$, from (22) we have $[\hat{\mathbf{w}}^*]_j<0$.
From Theorem 2, we have $j\notin\arg\min_{C\subseteq\hat{V}}\hat{F}(C)$. Noting that $A^*=\mathcal{E}\cup\arg\min_{C\subseteq\hat{V}}\hat{F}(C)$ and $j\notin\mathcal{E}$. Therefore $j\notin A^*$.
(iii) It is the immediate conclusion of (i) and (ii).
The proof is complete. □