Supplementary Document for "Submodularity Cuts and Applications"

In this supplementary material, we give the detail of a way to compute γ-extensions in Sect. A and an outline of the submodular maximization algorithm by Nemhauser and Wolsey [16] in Sect. B.

A Details of the computation of γ-extensions

Here, we describe a way to compute the γ-extensions of the directions $d_1, \ldots, d_n$ chosen in Sect. 3.1 of our manuscript. Let S be a set with $|S| = k$, and let i^* and j^* are specified elements in S and $V \setminus S$, respectively, and suppose that $S = \{i_1, \ldots, i_k\}$ with $i_1 = i^*$ and $V \setminus S = \{j_{k+1}, \ldots, j_n\}$ with $j_n = j^*$. Remember that we chose the directions $d_1, \ldots, d_n$ as

$$d_l = \begin{cases} e_{j^*} - e_{i_l} & \text{if } l \in \{1, \ldots, k\} \\ e_{j_l} - e_{j^*} & \text{if } l \in \{k+1, \ldots, n-1\} \\ -e_{j^*} & \text{if } l = n, \end{cases} \tag{8}$$

and gave the following lemma (Lemma 6 in Sect. 3.1):

Lemma 6 *For the directions $d_1, \ldots, d_n$ defined in Eqs. (8), a cone*

$$K(I_S; d_1, \ldots, d_n) = \{I_S + t_1 d_1 + \cdots + t_n d_n : t_l \geq 0\}$$

contains the polytope $D_0 = \{x \in \mathbb{R}^n : 0 \leq x_l \leq 1 \ (l = 1, \cdots, n), \ \sum_{l=1}^n x_l \leq k\}$.

Proof Let $T \subset V$ be an arbitrary fixed set with $|T| \leq k$. We have that $T \in S(D_0)$ and $I_T \in V(D_0)$. To show the assertion, it suffices to show $I_T \in K(I_S; d_1, \ldots, d_n)$ because of the one-to-one correspondence between $S(D_0)$ and $V(D_0)$. Let us partition the set S into tree parts $S_1 = \{i'_1, \ldots, i'_{|S \cap T|}\}$, $S_2 = \{i'_{|S \cap T|+1}, \ldots, i'_{|T|}\}$, and $S_3 = \{i'_{|T|+1}, \ldots, i'_k\}$, where

$$S_1 = S \cap T \quad \text{and} \quad S_2 \cup S_3 = S \setminus T.$$

In addition, we partition the set T into two parts $T_1 = T \cap S = S_1$ and $T_2 = T \setminus S = \{j'_{|S \cap T|+1}, \ldots, j'_{|T|}\}$. Then, we have

$$\begin{aligned} I_T &= I_S - I_{S \setminus T} + I_{T \setminus S} \\ &= I_S + (I_{T_2} - I_{S_2}) - I_{S_3} \\ &= I_S + \sum_{l=|S \cap T|+1}^{|T|} (e_{j'_l} - e_{i'_l}) + \sum_{l=|T|+1}^k (-e_{i'_l}) \\ &= I_S + \sum_{l=|S \cap T|+1}^{|T|} \{(e_{j'_l} - e_{j^*}) + (e_{j^*} - e_{i'_l})\} + \sum_{l=|T|+1}^k \{(e_{j^*} - e_{i'_l}) + (-e_{j^*})\}. \end{aligned}$$

Therefore, we obtain $I_T \in \{I_S + t_1 d_1 + \cdots + t_n d_n : t_l \geq 0\}$. ∎

Suppose that I_S is a vertex of a polytope $P \subseteq D_0$ and γ is a constant number such that $f(S) \leq \gamma$ and $f(S_{(i,j)}) \leq \gamma$ for any neighbor $S_{(i,j)}$ of S. For the directions $d_1, \ldots, d_n$ defined in Eqs. (8), the submodularity cuts algorithm requires the γ-extensions $y_l = I_S + \theta_l d_l$ $(l = 1, \ldots, n)$ with respect to the Lovász extension $\hat{f} : \mathbb{R}^n \to \mathbb{R}$. To obtain these γ-extensions, it suffices to consider the following three cases.

- For $i \in S$ and $j^* \in V \setminus S$, compute a γ-extension $y^{(1)} = I_S + \theta^{(1)}(e_{j^*} - e_i)$.
- For two distinct $j, j^* \in V \setminus S$, compute a γ-extension $y^{(2)} = I_S + \theta^{(2)}(e_j - e_{j^*})$.
- For $j^* \in V \setminus S$, compute a γ-extension $y^{(3)} = I_S + \theta^{(3)}(-e_{j^*})$.

Let us give explicit representations of $\theta^{(1)}, \theta^{(2)}$ and $\theta^{(3)}$.

A.1 Computation of $\theta^{(1)}$

We have $\theta^{(1)} = \max\{t \geq 0 : \hat{f}(y(t)) \leq \gamma\}$, where $y(t) := I_S + t(e_{j^*} - e_i)$. Note that $f(S) = \hat{f}(y(0)) \leq \gamma$ and $f(S_{(i,j^*)}) = \hat{f}(y(1)) \leq \gamma$. The convexity of $\hat{f}$ implies $\theta^{(1)} \geq 1$. Thus,

1
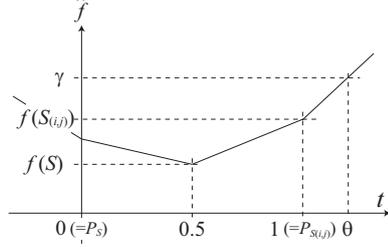
Figure 1: Determination of θ in computing the γ-extension through Lovász extension for $l \in \{1, \ldots, k\}$ (cf. Eq. (8)).

we consider a parameter $t \geq 1$. The different components of $\boldsymbol{y}(t)$ are ordered as $t \geq 1 > 0 \geq 1 - t$, and $\boldsymbol{y}(t)$ can be represented as

$$\boldsymbol{y}(t) = (t - 1) \cdot \boldsymbol{I}_{\{j^*\}} + 1 \cdot \boldsymbol{I}_{S_{(i,j^*)}} + (t - 1) \cdot \boldsymbol{I}_{V \setminus \{i\}} + (1 - t) \cdot \boldsymbol{I}_V.$$

In view of the definition of the Lovász extension, the value of $\hat{f}(\boldsymbol{y}(t))$ is given by

$$\hat{f}(\boldsymbol{y}(t)) = (t - 1) \cdot f(\{j^*\}) + 1 \cdot f(S_{(i,j^*)}) + (t - 1) \cdot f(V \setminus \{i\}) + (1 - t) \cdot f(V)$$

$$= f(S_{(i,j^*)}) + (t - 1)\Big\{ f(\{j^*\}) + f(V \setminus \{i\}) - f(V) \Big\}.$$

So, $\theta^{(1)}$ is bounded if and only if $\frac{\mathrm{d}}{\mathrm{d}t} \hat{f}(\boldsymbol{y}(t)) = f(\{j^*\}) + f(V \setminus \{i\}) - f(V) > 0$. As a result, we have the following (see also Fig. 1):

$$\theta^{(1)} = \begin{cases} 1 + \dfrac{\gamma - f(S_{(i,j^*)})}{f(\{j^*\}) + f(V \setminus \{i\}) - f(V)} & \text{if } f(\{j^*\}) + f(V \setminus \{i\}) - f(V) > 0, \\ +\infty & \text{otherwise.} \end{cases}$$

A.2 Computation of $\theta^{(2)}$

We have $\theta^{(2)} = \max\{t \geq 0 : \hat{f}(\boldsymbol{y}(t)) \leq \gamma\}$, where $\boldsymbol{y}(t) := \boldsymbol{I}_S + t(\boldsymbol{e}_j - \boldsymbol{e}_{j^*})$. Note that $f(S) = \hat{f}(\boldsymbol{y}(0)) \leq \gamma$. First, we suppose that $t \in [0, 1]$. The different components of $\boldsymbol{y}(t)$ are ordered as $1 \geq t > 0 \geq -t$ and $\boldsymbol{y}(t)$ can be represented as

$$\boldsymbol{y}(t) = (1 - t) \cdot \boldsymbol{I}_S + t \cdot \boldsymbol{I}_{S \cup \{j\}} + t \cdot \boldsymbol{I}_{V \setminus \{j^*\}} + (-t) \cdot \boldsymbol{I}_V.$$

Thus, the value of $\hat{f}(\boldsymbol{y}(t))$ is given by

$$\hat{f}(\boldsymbol{y}(t)) = (1 - t) \cdot f(S) + t \cdot f(S \cup \{j\}) + t \cdot f(V \setminus \{j^*\}) + (-t) \cdot f(V)$$

$$= f(S) + t \cdot \Big\{ f(S \cup \{j\}) + f(V \setminus \{j^*\}) - f(S) - f(V) \Big\} =: g^{\text{low}}(t).$$

If $\gamma < g^{\text{low}}(1)$, we obtain $\frac{\mathrm{d}}{\mathrm{d}t} g^{\text{low}}(t) = f(S \cup \{j\}) + f(V \setminus \{j^*\}) - f(S) - f(V) > 0$ and thus $\theta^{(2)} = \frac{\gamma - f(S)}{f(S \cup \{j\}) + f(V \setminus \{j^*\}) - f(S) - f(V)}$. Otherwise, we have $\theta^{(2)} \geq 1$.

Next, we suppose that $g^{\text{low}}(1) \leq \gamma$ and $1 \leq t$. The different components of $\boldsymbol{y}(t)$ are ordered as $t \geq 1 > 0 \geq -t$ and $\boldsymbol{y}(t)$ can be represented as

$$\boldsymbol{y}(t) = (t - 1) \cdot \boldsymbol{I}_{\{j\}} + 1 \cdot \boldsymbol{I}_{S \cup \{j\}} + t \cdot \boldsymbol{I}_{V \setminus \{j^*\}} + (-t) \cdot \boldsymbol{I}_V.$$

Thus, the value of $\hat{f}(\boldsymbol{y}(t))$ is given by

$$\hat{f}(\boldsymbol{y}(t)) = (t - 1) \cdot f(\{j\}) + 1 \cdot f(S \cup \{j\}) + t \cdot f(V \setminus \{j^*\}) + (-t) \cdot f(V)$$

$$= f(S \cup \{j\}) - f(\{j\}) + t \cdot \Big\{ f(\{j\}) + f(V \setminus \{j^*\}) - f(V) \Big\} =: g^{\text{high}}(t).$$

On the assumption that $g^{\text{low}}(1) \leq \gamma$, $\theta^{(2)}$ is bounded if and only if $\frac{\mathrm{d}}{\mathrm{d}t} g^{\text{high}}(t) = f(\{j\}) + f(V \setminus \{j^*\}) - f(V) > 0$. As a result, we obtain

$$\theta^{(2)} = \begin{cases} \dfrac{\gamma - f(S)}{f(S \cup \{j\}) + f(V \setminus \{j^*\}) - f(S) - f(V)} & \text{if } \gamma < g^{\text{low}}(1), \\ \dfrac{\gamma - f(S \cup \{j\}) + f(\{j\})}{f(\{j\}) + f(V \setminus \{j^*\}) - f(V)} & \text{if } \gamma \geq g^{\text{low}}(1) \text{ and } \frac{\mathrm{d}}{\mathrm{d}t} g^{\text{high}}(t) > 0, \\ +\infty & \text{if } \gamma \geq g^{\text{low}}(1) \text{ and } \frac{\mathrm{d}}{\mathrm{d}t} g^{\text{high}}(t) \leq 0. \end{cases}$$

Note that $g^{\text{low}}(1) = g^{\text{high}}(1)$, $\frac{\mathrm{d}}{\mathrm{d}t} g^{\text{low}}(t) \leq \frac{\mathrm{d}}{\mathrm{d}t} g^{\text{high}}(t)$ and $\hat{f}(\boldsymbol{y}(t)) = \min\{g^{\text{low}}(t), g^{\text{high}}(t)\}$ for all $t \geq 0$.

2

Algorithm 3 Description of the algorithm by Nemhauser and Wolsey.

1: Let $Q_1 = \{T_1, \ldots, T_m\}$ be a set of distinct subsets of V.
2: Set stop $\leftarrow$ false and $i \leftarrow 1$.
3: **while** stop = false **do**
4:     Solve the MIP problem (S1) with respect to Q_i, and let $(\eta_i, \boldsymbol{I}_{S_i})$ be an optimal solution.
5:     **if** $f(S_i) = \eta_i$ **then**
6:        stop $\leftarrow$ true (S_i is an optimal solution and η_i the optimal value).
7:     **else**
8:        Set $Q_{i+1} \leftarrow Q_i \cup \{S_i\}$ and $i \leftarrow i + 1$
9:     **end if**
10: **end while**

Remark: If $f(S) = \gamma$, then $\theta^{(2)}$ can be equal to 0. In such a case, we choose a small $\delta > 0$ and replace $\theta^{(2)}$ by δ so that the resulting submodularity cut H satisfies Lemma 4.

A.3 Computation of $\theta^{(3)}$

We have $\theta^{(3)} = \max\{t \geq 0 : \hat{f}(\boldsymbol{y}(t)) \leq \gamma\}$, where $\boldsymbol{y}(t) := \boldsymbol{I}_S - t \cdot \boldsymbol{e}_{j*}$. Note that $f(S) = \hat{f}(\boldsymbol{y}(0)) \leq \gamma$. Consider a parameter $t \geq 0$. The different components of $\boldsymbol{y}(t)$ are ordered as $1 > 0 \geq -t$ and $\boldsymbol{y}(t)$ can be represented as

$$\boldsymbol{y}(t) = 1 \cdot \boldsymbol{I}_S + t \cdot \boldsymbol{I}_{V \setminus \{j^*\}} + (-t) \cdot \boldsymbol{I}_V.$$

Thus, the value of $\hat{f}(\boldsymbol{y}(t))$ is given by

$$\hat{f}(\boldsymbol{y}(t)) = 1 \cdot f(S) + t \cdot f(V \setminus \{j^*\}) + (-t) \cdot f(V) = f(S) + t\Big\{f(V \setminus \{j^*\}) - f(V)\Big\}.$$

So, $\theta^{(3)}$ is bounded if and only if $\frac{\mathrm{d}}{\mathrm{d}t}\hat{f}(\boldsymbol{y}(t)) = f(V \setminus \{i\}) - f(V) > 0$. Thus, we have

$$\theta^{(3)} = \begin{cases} \dfrac{\gamma - f(S)}{f(V \setminus \{j^*\}) - f(V)} & \text{if } f(V \setminus \{i\}) - f(V) > 0, \\ +\infty & \text{otherwise.} \end{cases}$$

In particular, if f is nondecreasing, it holds that $f(V \setminus \{i\}) \leq f(V)$ and thus we have $\theta^{(3)} = +\infty$.

B   Outline of the algorithm by Nemhauser & Wolsey

Throughout this section, we assume that $f : 2^V \to \mathbb{R}$ is nondecreasing. Nemhauser and Wolsey [16] showed that the submodular maximization problem

$$\max_{S \subseteq V} f(S) \quad \text{s.t. } |S| \leq k \tag{1}$$

can be reformulated as

$$\begin{aligned} \max_{\eta, \boldsymbol{y}} \quad & \eta \\ \text{s.t.} \quad & \eta \leq f(S) + \sum_{j \in V \setminus S} \rho_j(S) y_j, \quad \forall S \in 2^V, \\ & \sum_{j \in V} y_j = k, \quad y_j \in \{0,1\}, \quad \forall j \in V, \end{aligned} \tag{10}$$

where $\rho_j(S) := f(S \cup \{j\}) - f(S)$. Specifically, $(\eta, \boldsymbol{y}) = (f(S), \boldsymbol{I}_S)$ is an optimal solution to (10) if and only if S is an optimal solution to (1). Let us denote the optimal value of (10) by η^*. The problem (10) is a mixed integer program (MIP) with $O(2^n)$ inequalities, and it would be intractable because of the number of constraints.

We describe the algorithm of Nemhauser and Wolsey [16] for the submodular maximization problem (1). Initially, we let $Q_1 = \{T_1, \ldots, T_m\}$ be a set of distinct subsets of V, and set $i \leftarrow 1$. In iteration $i \geq 1$, solve the following (relatively small) MIP

$$\begin{aligned} \max_{\eta, \boldsymbol{y}} \quad & \eta \\ \text{s.t.} \quad & \eta \leq f(S) + \sum_{j \in V \setminus S} \rho_j(S) y_j, \quad \forall S \in Q_i = \{T_1, \ldots, T_m, S_1, \ldots, S_{i-1}\}, \\ & \sum_{j \in V} y_j = k, \quad y_j \in \{0,1\}, \quad \forall j \in V, \end{aligned} \tag{S1}$$

3

and let $(\eta_i, \boldsymbol{I}_{S_i})$ be an optimal solution to (S1). Since the feasible region of (S1) is larger than or equal to that of (10), η_i is an upper bound on η^*. Thus, we have $f(S_i) \leq \eta^* \leq \eta_i$. Therefore, if $f(S_i) = \eta_i$, S_i is an optimal solution to the submodular maximization problem. If $f(S_i) < \eta_i$, set $Q_{i+1} \leftarrow Q_i \cup \{S_i\}$, $i \leftarrow i + 1$ and execute the next iteration. The description of the algorithm is given in Alg. 3. It terminates after at most $\binom{n}{k}$ iterations.

Note that $\eta_1 \geq \eta_2 \geq \cdots \geq \eta^*$. Thus, Alg. 3 can also be regarded as a subroutine just for the nonincreasing upper bound η_i.

4