Supplementary Material

1 Proof of Proposition 1

**Proposition 1.** Suppose \( f : 2^N \to \mathbb{R} \) is submodular. After each iteration of Algorithm 1 (Algorithm 2), we have \( X_{\min} \subseteq [X_1, Y_1] \) (\( X_{\max} \subseteq [X_1, Y_1] \)).

For Algorithm 1, the proof can be found in [3]. For Algorithm 2, the proof can be found in [2]. A proof using weaker assumption of quasi-submodular function is presented in [5]. We prove Proposition 1 here for completeness.

**Proof.** Algorithm 1. Obviously \( X_{\min} \subseteq [X_0, Y_0] \). Suppose \( X_{\min} \subseteq [X_k, Y_k] \), we now prove \( X_{\min} \subseteq [X_{k+1}, Y_{k+1}] \). Suppose \( X_\star \subseteq X_{\min} \) is a minimum of \( f \), then we have \( X_k \subseteq X_\star \subseteq Y_k \). For \( \forall i \in U_k \), if \( i \notin X_\star \), by submodularity, we have \( f(i|X_\star) \leq f(i|X_k) < 0 \), i.e., \( f(X_\star + i) < f(X_\star) \), which contradicts with the optimality of \( X_\star \). So we have \( U_k \subseteq X_\star \), and \( X_{k+1} = X_k \cup U_k \subseteq X_\star \). \( \forall j \in D_k \), if \( j \notin X_\star \), by submodularity, we have \( f(j|X_\star - j) \geq f(j|Y_k - j) > 0 \), i.e., \( f(X_\star) > f(X_\star - j) \), which also contradicts with the optimality of \( X_\star \). Therefore we have \( D_k \subseteq N \setminus X_\star \), and \( X_\star \subseteq Y_{k+1} = Y_k \setminus D_k \).

Now we have \( X_{k+1} \subseteq X_\star \subseteq Y_{k+1} \). Since \( X_\star \) can be an arbitrary element of \( X_{\min} \), we have \( X_{\min} \subseteq [X_{k+1}, Y_{k+1}] \).

**Algorithm 2.** Obviously \( X_{\max} \subseteq [X_0, Y_0] \). Suppose \( X_{\max} \subseteq [X_k, Y_k] \), we now prove \( X_{\max} \subseteq [X_{k+1}, Y_{k+1}] \). Suppose \( X_\star \subseteq X_{\max} \) is a maximum of \( f \), then we have \( X_k \subseteq X_\star \subseteq Y_k \). \( \forall i \in U_k \), if \( i \notin X_\star \), by submodularity, we have \( f(i|X_\star - i) \leq f(i|X_k) < 0 \), i.e., \( f(X_\star) < f(X_\star - i) \), which contradicts with the optimality of \( X_\star \). So we have \( U_k \subseteq N \setminus X_\star \), and \( X_\star \subseteq Y_{k+1} = Y_k \setminus U_k \). \( \forall j \in D_k \), if \( j \notin X_\star \), by submodularity, we have \( f(j|X_\star) \geq f(j|Y_k - j) > 0 \), i.e., \( f(X_\star + j) > f(X_\star) \), which also contradicts with the optimality of \( X_\star \). So we have \( D_k \subseteq X_\star \), and \( X_{k+1} = X_k \cup D_k \subseteq X_\star \).

Now we have \( X_{k+1} \subseteq X_\star \subseteq Y_{k+1} \). Since \( X_\star \) can be an arbitrary element of \( X_{\max} \), we have \( X_{\max} \subseteq [X_{k+1}, Y_{k+1}] \). \( \square \)

2 Reduction Rate of Algorithm 1

Figure 1 shows the reduction rates of Algorithm 1. All the settings are the same as those of Algorithm 2 in the paper.

3 More Experimental Results

3.1 Results of Maximization

In the paper we use the random bi-directional greedy method as the approximate solver for maximization. We also report the results of random permutation [3] and random local search [3]. The settings are the same as those in the paper. The results are shown in Figure 2 and Figure 3.

3.2 Results Using Real Data

Finally, we compare the results on real data. The objective function is the log-determinant function. For each test case, we randomly select 100 samples from the CIFAR dataset [4], and then we compute the similarity matrix as the positive definite matrix \( K \). Other settings are the same as those in the paper. The results are shown in Figure 4.

In Figure 4, the first three subfigures show the results of maximization using random local search, random permutation, and random bi-directional greedy, respectively. The last subfigure presents the results of minimization using the Fujishige-Wolfe minimum-norm point algorithm [1].

References


Figure 1: Average Reduction Rates of Minimization

Figure 2: Maximization Results Using Random Permutation [3]

Figure 3: Maximization Results Using Random Local Search [3]

Figure 4: Results of Log-determinant Function Using CIFAR Dataset