Supplementary Material

1 Proof of Proposition 1

Proposition 1. *Suppose $f : 2^N \mapsto \mathbb{R}$ is submodular. After each iteration of A1 (A2), we have $\mathcal{X}_{min} \subseteq [X_t, Y_t]$ ($\mathcal{X}_{max} \subseteq [X_t, Y_t]$).*

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 f can be found in [5]. We prove Proposition 1 here for completeness.

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

Now we have $X_{k+1} \subseteq X_* \subseteq Y_{k+1}$. Since X_* can be an arbitrary element of $\mathcal{X}_{min}$, we have $\mathcal{X}_{min} \subseteq [X_{k+1}, Y_{k+1}]$.

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

Now we have $X_{k+1} \subseteq X^* \subseteq Y_{k+1}$. Since X^* can be an arbitrary element of $\mathcal{X}_{max}$, we have $\mathcal{X}_{max} \subseteq [X_{k+1}, Y_{k+1}]$. □

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

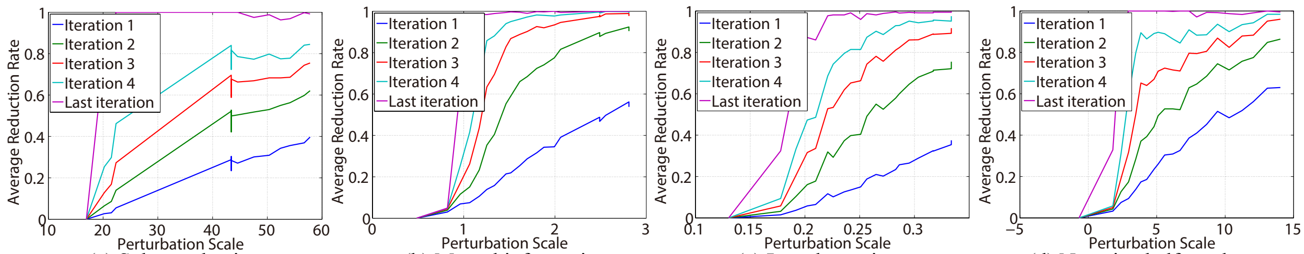
[1] Satoru Fujishige and Shigueo Isotani. A submodular function minimization algorithm based on the minimum-norm base. *Pacific Journal of Optimization*, 7(1):3–17, 2011.

[2] Boris Goldengorin. Maximization of submodular functions: Theory and enumeration algorithms. *European Journal of Operational Research*, 198(1):102–112, 2009.

[3] Rishabh Iyer, Stefanie Jegelka, and Jeff Bilmes. Fast semidifferential-based submodular function optimization. In *International Conference on Machine Learning*, pages 855–863, 2013.

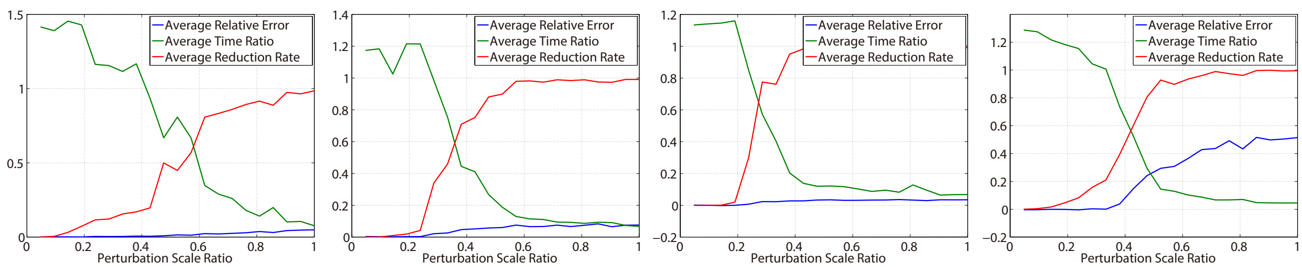
[4] Alex Krizhevsky. Learning multiple layers of features from tiny images, 2009.

[5] Jincheng Mei, Kang Zhao, and Bao-Liang Lu. On unconstrained quasi-submodular function optimization. In *AAAI Conference on Artificial Intelligence*, pages 1191–1197, 2015.



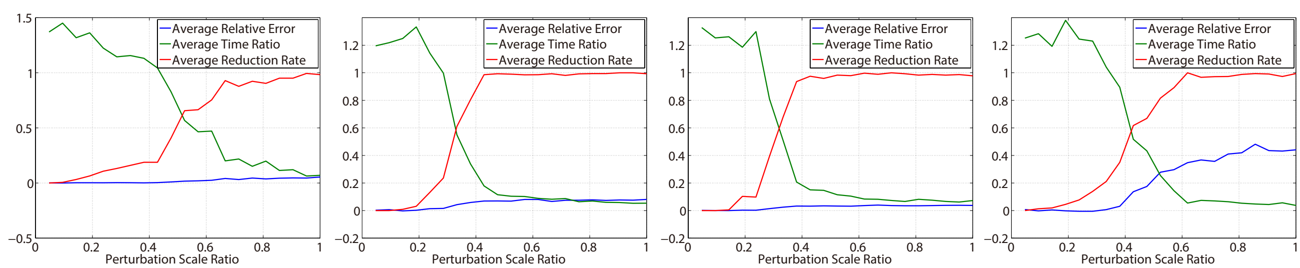
(a) Subset selection (b) Mutual information (c) Log-determinant (d) Negative half-products

Figure 1: Average Reduction Rates of Minimization



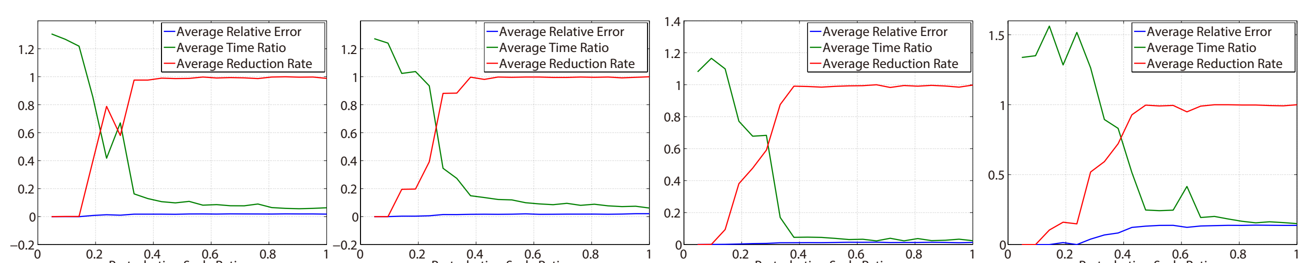
(a) Subset selection (b) Mutual information (c) Log-determinant (d) Negative half-products

Figure 2: Maximization Results Using Random Permutation [3]



(a) Subset selection (b) Mutual information (c) Log-determinant (d) Negative half-products

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



(a) Random local search (b) Random permutation (c) Random bi-directional greedy (d) Minimum-norm point

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