TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 07/20 pp1703–1708
DOI: 10.26599/TST.2023.9010107
Volume 29, Number 6, December 2024

Two-Stage Submodular Maximization Under Knapsack Problem

Zhicheng Liu, Jing Jin, Donglei Du, and Xiaoyan Zhang*

Abstract: Two-stage submodular maximization problem under cardinality constraint has been widely studied in machine learning and combinatorial optimization. In this paper, we consider knapsack constraint. In this problem, we give n articles and m categories, and the goal is to select a subset of articles that can maximize the function $F(S)$. Function $F(S)$ consists of m monotone submodular functions f_j, $j=1,2,\ldots,m$, and each f_j measures the similarity of each article in category j. We present a constant-approximation algorithm for this problem.

Key words: submodular function; knapsack constraint; matroid

1 Introduction

The problem we are interested in is related to the Combinatorial Representation Problem (CRP). In CRP, we are given a ground set V of n articles and a set N of m categories, and nonnegative monotone submodular functions $f_j: 2^V \to \mathbf{R}_{\geqslant 0}$ used to measure the similarity of each article in category j. The goal is to select a subset $S \subseteq V$ of articles that best represent the different categories. This problem can be formulated as the following two-stage submodular maximization problem according to Ref. [2]:

$$\max_{S \in C_1} F(S) := \sum_{j=1}^{m} \max_{T \in C_2} f_j(T) \tag{1}$$

where $C_1 \subseteq V$ and $C_2 \subseteq S$ are two constraint sets, and f_j is a non-negative monotone submodular function. For example, C_1 may be a cardinality constraint and C_2 may be a matroid constraint[2–4]. The function $f: 2^V \to \mathbf{R}$ is said to be submodular if $f(X)+f(Y) \geqslant f(X \cap Y)+f(X \cup Y)$. That is $f(e \mid X) \geqslant f(e \mid Y)$ for $X \subseteq Y \subset V$ and $e \in V \setminus Y$, where $f(e \mid X) = f(e \cup X) - f(X)$. A set function f is called monotone if $f(X) \leqslant f(Y)$ for all $X \subseteq Y \subseteq V$, and it is said to be normalized when $f(\varnothing)=0$. There are several papers considering the related problems. In Ref. [2], the authors used local search to design an approximation algorithm and get a $\frac{1}{2}\left(1-\mathrm{e}^{-1}\right)$-approximation ratio. Recently, the authors in Ref. [4] used the replacement greedy algorithm to achieve an improved approximation ratio $\frac{1}{2}\left(1-\mathrm{e}^{-2}\right)$, and the authors in Ref. [3] developed the first streaming and distributed algorithms for this problem. In addition, authors in Ref. [5] considered the generalized k-matroids constraint. Based on generalized submodularity ratio[6], authors in Ref. [7] developed a parameterized streaming algorithm for the two-stage submodular maximization. In Ref. [8], authors considered two-stage submodular maximization based on submodularity curvature[9–11].

• Zhicheng Liu is with Beijing Institute for Scientific and Engineering Computing, Beijing University of Technology, Beijing 100124, China. E-mail: manlzhic@emails.bjut.edu.cn.
• Jing Jin is with College of Taizhou, Nanjing Normal University, Taizhou 225300, China. E-mail: jinjing19841@126.com.
• Donglei Du is with Faculty of Management, University of New Brunswick, Fredericton, E3B 5A3, Canada. E-mail: ddu@unb.ca.
• Xiaoyan Zhang is with School of Mathematical Science, Institute of Mathematics, and Ministry of Education Key Laboratory of NSLSCS, Nanjing Normal University, Nanjing 210023, China. E-mail: zhangxiaoyan@njnu.edu.cn.
• A preliminary version of this paper appears in Proceedings of the 17th Annual Conference on Theory and Applications of Models of Computation (TAMC 2022)[1].
* To whom correspondence should be addressed.
Manuscript received: 2022-12-15; revised: 2023-03-08; accepted: 2023-09-17

In Ref. [12], authors considered the case when the objective can be negative and nonmonotone. To the best of our knowledge, there are no previous results on C_1 being a knapsack constraint which can be seen as an extension of the cardinality constraint.

In this paper, we assume that C_1 is a knapsack constraint and $\mathcal{I}(\mathcal{S})$ is the family of the common independent sets of a k-matroid over the same ground set $S \subseteq V$. Our contribution is to present a $\frac{1}{2(k+1)}\left(1-\mathrm{e}^{-(k+1)}\right)$-approximation algorithm for this problem with a new analysis.

The rest of the paper is organized as follows. In Section 2 we introduce some definitions and properties of submodular function and matroid. In Section 3, we present the algorithms and analysis of Eq. (1). Finally, we offer concluding remarks in Section 4.

2 Preliminary

Definition 1 Given a ground set $V=\{1, 2, \ldots, n\}$ and a family of subsets $\mathcal{I}$ of V, a matroid $\mathcal{M}=(V, \mathcal{I})$ satisfies the following properties:

(1) $\varnothing \in \mathcal{I}$;

(2) If $A \subseteq B \in \mathcal{I}$, then $A \in \mathcal{I}$;

(3) If $A, B \in \mathcal{I}$ and $|A| < |B|$, then there exists an element $u \in B \backslash A$ for which $A \cup \{u\} \in \mathcal{I}$.

Next, we introduce some properties of matroid and submodular function. We will need the following matroid property from Ref. [13] later.

Let $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k$ be k arbitrary matroids on the common ground set V. For each matroid $\mathcal{M}_j$ (with $j \in \{1, 2, \ldots, k\}$) we denote the set of its independent sets by $\mathcal{I}_j$.

Proposition 1 Let $\mathcal{M}_j=(V, \mathcal{I}_j)$ be a matroid for every $j \in \{1, 2, \ldots, k\}$. For any two independent sets $A, B \in \mathcal{I}_j$, there exists a mapping π_j: $B \setminus A \to A \setminus B \cup \{\varnothing\}$, such that

(1) $(A \backslash \pi_j(b)) \cup b \in \mathcal{I}_j$ for all $b \in B \backslash A$;

(2) $\left|\pi_j^{-1}(a)\right| \leqslant 1$ for all $a \in A \backslash B$;

(3) let $A_b=\{\pi_1(b), \pi_2(b), \ldots, \pi_k(b)\}$, then $(A \backslash A_b) \cup b \in \cap_{j=1}^{k} \mathcal{I}_j$ for all $b \in B \backslash A$.

Proposition 2 For any submodular function f: $2^V \to \mathbf{R}_+$ and $X, Y \subseteq V$, we have

$$\sum_{u \in X} (f(Y \cup \{u\}) - f(Y)) \geqslant f(X \cup Y) - f(Y).$$

The next property is from Ref. [14].

Proposition 3 Consider a monotone submodular function f: $2^V \to \mathbf{R}_+$. Let $X, Y \subseteq V$, and $\{T_i\}_{i=1}^{\ell}$ be a collection of subsets of $Y \setminus X$, such that each element of $Y \setminus X$ appears in at most k of the subsets. Then

$$\sum_{i=1}^{\ell} (f(Y) - f(Y \setminus T_i)) \leqslant k (f(Y) - f(Y \cap X)).$$

3 Two-Stage Submodular Maximization Subject to Knapsack and Matroid Constraints

We consider Eq. (1) by offering an approximation algorithm along with its analysis in Sections 3.1 and 3.2, respectively.

3.1 Algorithm

Denote the gain of adding set X to the set A as follows:

$$\Delta_j^f(X, A) = f_j(X \cup A) - f_j(A).$$

Denote the gain of removing a set $Y \subseteq A$ and replacing it with element x as follows:

$$\nabla_j^f(x, Y, A) = f_j(\{x\} \cup A \setminus Y) - f_j(A).$$

Consider the set A and define the set $\mathcal{I}(x, A) = \{Y \subseteq A : A \cup \{x\} \setminus Y \in \mathcal{I}\}$. Define the replacement gain of x as follow:

$$\nabla_j^f(x, A) = \begin{cases} \Delta_j^f(x, A), & \text{if } A \cup \{x\} \in \mathcal{I}; \\ \max\left\{0, \max\limits_{Y \in I(x, A)} \nabla_j^f(x, Y, A)\right\}, & \text{otherwise.} \end{cases}$$

Let $\mathrm{Rep}_j^f(x, A)$ be the set that is replaced by x,

$$\mathrm{Rep}_j^f(x, A) = \begin{cases} \varnothing, & \text{if } A \cup \{x\} \in \mathcal{I}; \\ \arg\max\limits_{Y \in I(x, A)} \nabla_j^f(x, Y, A), & \text{otherwise.} \end{cases}$$

The algorithm starts with an empty set $S=\varnothing$, and chooses an element with the largest ratio of marginal gain over cost in every round. Property 1 guarantees the correctness of the our algorithm.

3.2 Analysis

Define S^* as the optimal solution of Eq. (1),

$$S^* = \arg\max_{c(S) \leqslant B} \sum_{j=1}^{m} \max_{T \in I(S)} f_j(T).$$

Denote S_j^* as the optimal solution of f_j,

$$S_j^* = \arg\max_{T \in I(S^*)} f_j(T).$$

Based on Algorithm 1, we introduce the following notations:

Algorithm 1 Replacement modified greedy

1: $U \leftarrow V, S^F \leftarrow \varnothing, S \leftarrow \varnothing, T_j^F \leftarrow \varnothing, T_j \leftarrow \varnothing \ (\forall 1 \leqslant j \leqslant m)$
2: **while** $U \neq \varnothing$ **do**
3: $\quad p \leftarrow \arg\max_{e \in U} \frac{\sum_{j=1}^m \nabla_j^f(e, T_j)}{c_e}$;
4: $\quad U \leftarrow U \setminus \{p\}$;
5: $\quad$ **if** $\sum_{u \in S} c_u + c_p \leqslant B$ **then**
6: $\quad\quad S \leftarrow S \cup \{p\}$;
7: $\quad\quad$ **for** $1 \leqslant j \leqslant m$ **do**
8: $\quad\quad\quad$ **if** $\nabla_j^f(p, T_j) > 0$ **then**
9: $\quad\quad\quad\quad T_j \leftarrow T_j \cup \{p\} \setminus \mathrm{Rep}_j^f(p, T_j)$;
10: $\quad\quad\quad$ **end if**
11: $\quad\quad$ **end for**
12: $\quad$ **end if**
13: **end while**
14: $u^* \leftarrow \arg\max_{u \in V, c_u \leqslant B} \sum_{j=1}^m f_j(u^*)$;
15: **if** $\sum_{j=1}^m f_j(T_j) > \sum_{j=1}^m f_j(u^*)$ **then**
16: $\quad S^F \leftarrow S, T_1^F \leftarrow T_1, T_2^F \leftarrow T_2, \ldots, T_m^F \leftarrow T_m$;
17: **else**
18: $\quad S^F \leftarrow u^*, T_1^F \leftarrow u^*, T_2^F \leftarrow u^*, \ldots, T_m^F \leftarrow u^*$;
19: **end if**

- S is the solution obtained by the greedy heuristic;
- v_i is the i-th unit added to S ($i = 1, 2, \ldots, |S|$);
- S_i is the set of function $F(S)$ obtained by greedy algorithm after adding v_i (i.e., $S_i = \cup_{k=1}^i \{v_k\}$, for $i = 1, 2, \ldots, |S|$, with $S_0 = \varnothing$, $S_{|S|} = S$); and
- T_j^i is the set which is chosen by $f_j(T)$ after adding v_i to the set S.

We first establish the following two Lemmas to bound the increment in each iteration.

Lemma 1 For $i = 1, 2, \ldots, |S|+1$, we have

$$\sum_{e \in S^*} \sum_{j=1}^m \nabla_j^f(e, T_j^{i-1}) \leqslant \frac{B}{c_{v_i}} \sum_{j=1}^m \nabla_j^f(v_i, T_j^{i-1}).$$

Proof From Line 3 of Algorithm 1, we have

$$\frac{\sum_{j=1}^m \nabla_j^f(e, T_j^{i-1})}{c_e} \leqslant \frac{\sum_{j=1}^m \nabla_j^f(v_i, T_j^{i-1})}{c_{v_i}}, \ \forall e \in S^*.$$

Thus,

$$\sum_{e \in S^*} \sum_{j=1}^m \nabla_j^f(e, T_j^{i-1}) \leqslant \sum_{j=1}^m \nabla_j^f(v_i, T_j^{i-1}) \frac{\sum_{e \in S^*} c_e}{c_{v_i}} \leqslant \frac{B}{c_{v_i}} \sum_{j=1}^m \nabla_j^f(v_i, T_j^{i-1}).$$

■

Lemma 2 For $i = 1, 2, \ldots, |S|+1$, we have

$$\frac{B}{c_{v_i}} \sum_{j=1}^m \nabla_j^f(v_i, T_j^{i-1}) \geqslant \sum_{j=1}^m \sum_{e \in S_j^* \setminus T_j^{i-1}} (\Delta_j^f(e, T_j^{i-1}) - \Delta_j^f(A_e, \{e\} \cup T_j^{i-1} \setminus A_e)).$$

Proof Lemma 1 implies that

$$\sum_{e \in S^*} \sum_{j=1}^m \nabla_j^f(e, T_j^{i-1}) \leqslant \frac{B}{c_{v_i}} \sum_{j=1}^m \nabla_j^f(v_i, T_j^{i-1}).$$

From Property 1, there exist mappings $\pi_t : S_j^* \setminus T_j^{i-1} \to T_j^{i-1} \setminus S_j^* \cup \{\varnothing\}$ ($t \in \{1, 2, \ldots, k\}$), such that $(T_j^{i-1} \setminus A_e) \cup \{e\} \in \cap_{t=1}^k \mathcal{I}_t$, where $A_e = \{\pi_1(e), \pi_2(e), \ldots, \pi_k(e)\}$. Therefore,

$$\begin{aligned}
&\sum_{e \in S^*} \sum_{j=1}^m \nabla_j^f(e, T_j^{i-1}) = \sum_{j=1}^m \sum_{e \in S^*} \nabla_j^f(e, T_j^{i-1}) \geqslant \\
&\sum_{j=1}^m \sum_{e \in S_j^* \setminus T_j^{i-1}} \nabla_j^f(e, T_j^{i-1}) \geqslant \\
&\sum_{j=1}^m \sum_{e \in S_j^* \setminus T_j^{i-1}} \left(f_j(\{e\} \cup T_j^{i-1} \setminus A_e) - f_j(T_j^{i-1})\right) = \\
&\sum_{j=1}^m \sum_{e \in S_j^* \setminus T_j^{i-1}} (f_j(\{e\} \cup T_j^{i-1} \setminus A_e) - \\
&\quad f_j(\{e\} \cup T_j^{i-1}) + f_j(\{e\} \cup T_j^{i-1}) - f_j(T_j^{i-1})) = \\
&\sum_{j=1}^m \sum_{e \in S_j^* \setminus T_j^{i-1}} \left(\Delta_j^f(e, T_j^{i-1}) - (f_j(\{e\} \cup T_j^{i-1}) - \right. \\
&\quad \left. f_j(\{e\} \cup T_j^{i-1} \setminus A_e))\right) = \sum_{j=1}^m \sum_{e \in S_j^* \setminus T_j^{i-1}} (\Delta_j^f(e, T_j^{i-1}) - \\
&\quad \Delta_j^f(A_e, \{e\} \cup T_j^{i-1} \setminus A_e)),
\end{aligned}$$

where the first inequality follows because $\nabla_j^f(e, T_j^{i-1}) \geqslant 0$ and $S_j^* \setminus T_j^{i-1} \subseteq S^*$, and the second is due to Property 1 and the definition of $\nabla_j^f(e, T_j^{i-1})$. ■

Lemma 3 For $i = 1, 2, \ldots, |S|+1$, we have $\sum_{j=1}^m \sum_{e \in S_j^* \setminus T_j^{i-1}} \left(\Delta_j^f(e, T_j^{i-1}) - \Delta_j^f(A_e, \{e\} \cup T_j^{i-1} \setminus A_e)\right) \geqslant \sum_{j=1}^m \left(f_j(S_j^*) - (k+1) f_j(T_j^{i-1})\right)$.

Proof Denote $A_e^r = \{\pi_1(e), \pi_2(e), \ldots, \pi_r(e)\}$, for $r = 1, 2, \ldots, k$ and $A_e^0 = \varnothing$, $A_e^k = A_e$. We have

$$\Delta_j^f(A_e,\{e\}\cup T_j^{i-1}\setminus A_e)=$$
$$f_j(\{e\}\cup T_j^{i-1})-f_j(\{e\}\cup T_j^{i-1}\setminus A_e)=$$
$$f_j(\{e\}\cup T_j^{i-1})-f_j(\{e\}\cup T_j^{i-1}\setminus\{\pi_1(e),\pi_2(e),\ldots,\pi_k(e)\})=$$
$$\sum_{r=1}^{k}(f_j(\{e\}\cup T_j^{i-1}\setminus A_e^{r-1})-f_j(\{e\}\cup T_j^{i-1}\setminus A_e^{r}))\leqslant$$
$$\sum_{r=1}^{k}\left(f_j(T_j^{i-1}\setminus A_e^{r-1})-f_j(T_j^{i-1}\setminus A_e^{r})\right)=$$
$$f_j(T_j^{i-1})-f_j(T_j^{i-1}\setminus A_e)=\Delta_j^f(A_e,T_j^{i-1}\setminus A_e),$$

where the inequality is from the submodularity of f_j.

So we have

$$\sum_{j=1}^{m}\sum_{e\in S_j^*\setminus T_j^{i-1}}\left(\Delta_j^f(e,T_j^{i-1})-\Delta_j^f(A_e,\{e\}\cup T_j^{i-1}\setminus A_e)\right)\geqslant$$
$$\sum_{j=1}^{m}\sum_{e\in S_j^*\setminus T_j^{i-1}}\left(\Delta_j^f(e,T_j^{i-1})-\Delta_j^f(A_e,T_j^{i-1}\setminus A_e)\right).$$

Property 2 implies that

$$\sum_{e\in S_j^*\setminus T_j^{i-1}}\Delta_j^f(e,T_j^{i-1})=$$
$$\sum_{e\in S_j^*\setminus T_j^{i-1}}\left(f_j(\{e\}\cup T_j^{i-1})-f_j(T_j^{i-1})\right)\geqslant$$
$$f_j(S_j^*\cup T_j^{i-1})-f_j(T_j^{i-1}).$$

Property 3 implies that

$$\sum_{e\in S_j^*\setminus T_j^{i-1}}\Delta_j^f(A_e,T_j^{i-1}\setminus A_e)=$$
$$\sum_{e\in S_j^*\setminus T_j^{i-1}}\left(f_j(T_j^{i-1})-f_j(T_j^{i-1}\setminus A_e)\right)\leqslant$$
$$k\left(f_j(T_j^{i-1})-f_j(T_j^{i-1}\cap S_j^*)\right)\leqslant kf_j(T_j^{i-1}).$$

Together, we have

$$\sum_{j=1}^{m}\sum_{e\in S_j^*\setminus T_j^{i-1}}\left(\Delta_j^f(e,T_j^{i-1})-\Delta_j^f(A_e,\{e\}\cup T_j^{i-1}\setminus A_e)\right)\geqslant$$
$$\sum_{j=1}^{m}\left(f_j(S_j^*\cup T_j^{i-1})-(k+1)f_j(T_j^{i-1})\right)\geqslant$$
$$\sum_{j=1}^{m}\left(f_j(S_j^*)-(k+1)f_j(T_j^{i-1})\right).$$

■

According to Lemmas 2 and 3, we have the following corollary.

Corollary 1 For $i=1,2,\ldots,|S|+1$, we have

$$\frac{B}{c_{v_i}}\sum_{j=1}^{m}\nabla_j^f(v_i,T_j^{i-1})\geqslant\sum_{j=1}^{m}\left(f_j(S_j^*)-(k+1)f_j(T_j^{i-1})\right).$$

For convenience, we denote $X_{i-1}=\sum_{j=1}^{m}f_j(T_j^{i-1})$ and $X^*=\sum_{j=1}^{m}f_j(S_j^*)$. According to the Corollary 1, we have

$$X_i-X_{i-1}\geqslant\frac{c_{v_i}}{B}(X^*-(k+1)X_{i-1})\tag{2}$$

Lemma 4 $\forall i=1,2,\ldots,|S|+1$, if we assume $X^*\geqslant(k+1)X_i$, then we have

$$(k+1)c_{v_i}\leqslant B,\ \ \forall i=1,2,\ldots,|S|+1.$$

Proof Suppose for contradiction that there exists $j\leqslant|S|+1$, such that $(k+1)c_{v_j}>B$, then

$$X_j\geqslant\frac{c_{v_j}}{B}(X^*-(k+1)X_{j-1})+X_{j-1}>$$
$$\frac{1}{k+1}(X^*-(k+1)X_{j-1})+X_{j-1}=\frac{1}{k+1}X^*,$$

which contradicts the assumption. ■

Theorem 1 Algorithm 1 returns a set S_F, such that

$$F(S_F)\geqslant\frac{1}{2(k+1)}\left(1-\mathrm{e}^{-(k+1)}\right)F(S^*).$$

Proof We consider two cases.

Case 1 if there exists t, such that $X^*<(k+1)X_t$, then

$$F(S_F)\geqslant F(X_t)\geqslant\frac{1}{k+1}F(S^*).$$

Case 2 $\forall i=1,2,\ldots,|S|+1$, we have

$$X^*\geqslant(k+1)X_i.$$

Rearranging Inequality (2), we obtain

$$\frac{\frac{1}{k+1}X^*-X_i}{\frac{1}{k+1}X^*-X_{i-1}}\leqslant1-\frac{(k+1)c_{v_i}}{B}.$$

Therefore,

$$\frac{1}{k+1}X^*-X_{|S|+1}\leqslant\prod_{i=1}^{|S|+1}\left(1-\frac{(k+1)c_{v_i}}{B}\right)\frac{1}{k+1}X^*\leqslant$$
$$\prod_{i=1}^{|S|+1}e^{-\frac{(k+1)c_{v_i}}{B}}\frac{1}{k+1}X^*=e^{-\frac{\sum_{i=1}^{|S|+1}(k+1)c_{v_i}}{B}}\frac{1}{k+1}X^*\leqslant$$
$$e^{-\frac{(k+1)B}{B}}\frac{1}{k+1}X^*=e^{-(k+1)}\frac{1}{k+1}X^*\tag{3}$$

which is equivalent to

$$X_{|S|+1}\geqslant\frac{1}{k+1}\left(1-e^{-(k+1)}\right)X^*,$$

where the first inequality in Formula (3) follows that $X_0=0$, the second inequality in Formula (3) holds because $1-x \leqslant e^{-x}$, and the third inequality in Formula (3) is due to $\sum_{i=1}^{|S|+1} c_{v_i} > B$.

From Algorithm 1, we also have

$$X_{|S|+1} - X_{|S|} \leqslant \sum_{j=1}^{m} \left(f_j(T_j^{|S|} \cup \{v_{|S|+1}\}) - f_j(T_j^{|S|}) \right) \leqslant$$
$$\sum_{j=1}^{m} \left(f_j(\{v_{|S|+1}\}) - f_j(\varnothing) \right) = \sum_{j=1}^{m} \left(f_j(\{v_{|S|+1}\}) \right) \leqslant \sum_{j=1}^{m} f_j(\{u^*\}),$$

where the first inequality is due to Line 14 of Algorithm 1 and the second inequality follows from the submodularity of f_j.

Hence,

$$\sum_{j=1}^{m} f_j(\{u^*\}) + X_{|S|} \geqslant X_{|S|+1} \geqslant \frac{1}{k+1}\left(1 - e^{-(k+1)}\right) X^*,$$

implying that

$$\max\left\{ \sum_{j=1}^{m} f_j(\{u^*\}), X_{|S|} \right\} \geqslant \frac{1}{2(k+1)}\left(1 - e^{-(k+1)}\right) X^*. \qquad \blacksquare$$

4 Conclusion

In this paper, we consider two-stage submodular maximization problem under knapsack constraint which can be seen as a generalizati on of cardinality constraint, and we present a constant approximation algorithm for this problem with a new analysis.

Acknowledgment

The research was supported by the National Natural Science Foundation of China (Nos. 12131003, 12271259, 11371001, 11771386, and 11728104), the Natural Sciences and Engineering Research Council of Canada (NSERC) (No. 06446), the Natural Science Foundation of Jiangsu Province (No. BK20200267), and Qinglan Project.

References

[1] Z. Liu, J. Jin, D. Du, and X. Zhang, Two-stage submodular maximization under knapsack and matroid constraints, in *Proc. of 17th Annual Conf. on Theory and Applications of Models of Computation* (*TAMC*), Tianjin, China, pp.140-154, 2022.

[2] E. Balkanski, B. Mirzasoleiman, A. Krause, and Y. Singer, Learning sparse combinatorial representations via two-stage submodular maximization, in *Proc. of The 33rd Int. Conf. on Machine Learning*, New York, NY, USA, pp. 2207–2216, 2016.

[3] M. Mitrovic, E. Kazemi, M. Zadimoghaddam, and A. Karbasi, Data summarization at scale: A two-stage submodular approach, arXiv preprint arXiv: 1806.02815, 2018.

[4] S. Stan, M. Zadimoghaddam, A. Krause, and A. Karbasi, Probabilistic submodular maximization in sub-linear time, in *Proc. of the 34th Int. Conf. on Machine Learning,* Sydney, Australia, 2017, pp. 3241−3250.

[5] R. Yang, S. Gu, C. Gao, W. Wu, H. Wang, and D. Xu, A constrained two-stage submodular maximization, *Theor. Comput. Sci.*, vol. 853, pp. 57–64, 2021.

[6] S. Gong, Q. Nong, W. Liu, and Q. Fang, Parametric monotone function maximization with matroid constraints, *J. Glob. Optim.*, vol. 75, no. 3, pp. 833–849, 2019.

[7] R. Yang, D. Xu, L. Guo, and D. Zhang, Parametric streaming two-stage submodular maximization, in *Lecture Notes in Computer Science*, G. Goos and J. Hartmanis, eds. Cham, Switzerland: Springer International Publishing, pp. 193–204.

[8] Y. Li, Z. Liu, C. Xu, P. Li, X. Zhang, and H. Chang, Two-stage submodular maximization under curvature, *J. Comb. Optim.*, vol. 45, no. 2, p. 77, 2023.

[9] M. Conforti and G. Cornuéjols, Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the Rado-Edmonds theorem, *Discrete Appl. Math.*, vol. 7, no. 3, pp. 251–274, 1984.

[10] M. Sviridenko, J. Vondrák, and J. Ward, Optimal approximation for submodular and supermodular optimization with bounded curvature, *Math. Oper. Res.*, vol. 42, no. 4, pp. 1197–1218, 2017.

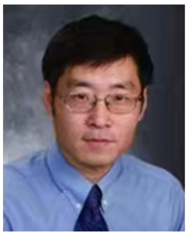
[11] Y. Yoshida, Maximizing a monotone submodular function with a bounded curvature under a knapsack constraint, *SIAM J. Discrete Math.*, vol. 33, no. 3, pp. 1452–1471, 2019.

[12] Z. Liu, H. Chang, R. Ma, D. Du, and X. Zhang, Two-stage submodular maximization problem beyond non-negative and monotone, *Lecture Notes in Computer Science*. Cham Switzerland: Springer International Publishing, 2020, pp. 144–155.

[13] J. Lee, V. S. Mirrokni, V. Nagarajan, and M. Sviridenko, Maximizing nonmonotone submodular functions under matroid or knapsack constraints, *SIAM J. Discrete Math.*, vol. 23, no. 4, pp. 2053–2078, 2010.

[14] J. Ward, Oblivious and non-oblivious local search for combinatorial optimization, PhD thesis, University of Toronto, Canada, 2012.

Zhicheng Liu received the MS degree from Nanjing Normal University, China in 2020. He is a PhD candidate at Beijing Institute for Scientifific and Engineering Computing, Beijing University of Technology, China. His research areas include operations research, combinatorial optimization, and approximations algorithms.



Donglei Du is a professor in quantitative method and quantitative investment management at Faculty of Management, University of New Brunswick, Canada. He received the PhD degree in computer science from University of Texas, USA in 2003, and another PhD degree in operations research from Institute of Applied Mathematics, Chinese Academy of Science in 1996. His main research interests are quantitative investment management, quantitative finance, FinTech (blockchain and cryptocurrencies) operations research, algorithmic trading, combinatorial optimization, approximations algorithms, robust optimization, social network analysis, algorithmic game theory, supply chain management, facility location, and machine scheduling. His publications have appeared in top-tier journals, including *Operations Research*, *Algorithmica*, *SIAM Journal on Discrete Mathematics*, *European Journal of Operation Research*, and *Omega*.

Jing Jin received the PhD degree in operation research from Nanjing Normal University, China in 2018. She is currently a full professor at College of Taizhou, Nanjing Normal University, China. Her major research interests include algorithm design, computational complexity analysis, graph theory, number theory, as well as combinatorial optimization and operation research.

Xiaoyan Zhang received the first PhD degree in applied mathematics from Nankai University, China in 2006, and the second PhD degree in computer science from University of Twente, the Netherlands in 2014. He is currently a full professor at Nanjing Normal University, China. He has published more than 70 academic papers in reputable journals, such as *SIAM Journal on Computing* (*SICOMP*), *SIAM Journal on Scientific Computing* (*SISC*), *SIAM Journal on Discrete Mathematics* (*SIDMA*), *Journal of Graph Theory* (*JGT*), and *IEEE Transactions on Information Theory* (*TIT*). His major research interests include efficient algorithm design and computational complexity analysis, particularly for combinatorial optimization, graph algorithms and networks, and VLSI.