Frank Wolfe Algorithm for Nonmonotone One-sided Smooth Function Maximization Problem

Hongxiang Zhang, Chunlin Hao, Wenying Guo and Yapu Zhang

November 4, 2022

Frank Wolfe Algorithm for Nonmonotone One-sided Smooth Function Maximization Problem

Hongxiang Zhang[1], Chunlin Hao[1], Wenying Guo[2], Yapu Zhang[1⋆]

[1] Institute of Operations Research and Information Engineering, Beijing University of Technology, Beijing 100124, P.R. China.
[2] School of Statistics, Capital University of Economics and Business, Beijing 100070, P.R. China.
zhanghx010@emails.bjut.edu.cn, haochl@bjut.edu.cn, guowy@cueb.edu.cn, zhangyapu@bjut.edu.cn

Abstract. In this paper, we study the problem of maximizing a nonmonotone one-sided-η smooth (OSS for short) function $\psi(x)$ under a downwards-closed convex polytope constraint. The concept of OSS was first proposed by Mehrdad et al. [1, 2] to express the properties of multilinear extension of some set functions. It is a generalization of the continuous DR submodular function. The OSS property guarantees an alternative bound based on Taylor expansion. If the objective function is nonmonotone diminishing return (DR) submodular, Bian et al. [3] gave a $1/e$ approximation algorithm with a regret bound $O(\frac{LD^2}{2K})$. On general convex sets, Dürr et al. [4] gave a $\frac{1}{3\sqrt{3}}$ approximation solution with $O(\frac{LD^2}{(\ln K)^2})$ regrets. In this paper, we consider maximizing the more general OSS function, and by adjusting the iterative step of the Jump-Start Frank Wolfe algorithm, an approximation of $1/e$ can still be obtained in the case of a larger regret bound $O(\frac{L(\mu D)^2}{2K})$. (where L, μ, D are some parameters, see Table 1). The larger the parameter η we choose, the more regrets we will receive, because of $\mu = \left(\frac{\beta}{\beta+1}\right)^{-2\eta}$ $(\beta \in (0, 1])$.

Keywords: approximation algorithm · one-sided smooth · nonmonotone · Frank Wolfe

1 Introduction

The concept of one-sided-smooth OSS was first proposed by Mehrdad et al. [1, 2] to express the properties of multilinear continuous extension of set functions or nonconvex functions. For example, the diminishing return (DR) submodular function is a case where $\eta = 0$, the multilinear extension [5] of proportionally submodular funtions is the case of $\eta = 1$, the multilinear extension of the diversity functions is a more general case ($\eta = 1, 2, 3...$) [1]. One-sided η-smooth functions have an important position in many fields, such as web search [6], machine learning [7, 8], document aggregation [9], recommender systems [10, 11].

⋆ Corresponding author.

A function ψ is one-sided-smooth if

$$\frac{1}{2}v^T\nabla^2\psi(x)v \leq \eta \cdot \frac{\|v\|_1}{\|x\|_1}v^T\nabla\psi(x)$$

for all $x, v \geq 0, x \neq 0$. In this paper, we discuss a nonmonotone OSS function maximization problem:

$$\max \psi(x) \quad s.t.\ x \in [0,1]^n, x \in \mathbb{P} \tag{1}$$

where $\mathbb{P}$ is a downwards-closed convex polytope, and it has a upper bound vector ϖ (i.e. any $x \in \mathbb{P}, x \leq \varpi$). $\psi : [0,1]^n \to R_+$ is a nonnegative nonmonotone normalized OSS function. Next, we define the regret function for the OSS maximization problem

$$r \cdot \max_{x\in\mathbb{P}} \psi(x) - \psi(x_K) \leq \pi(K), \tag{2}$$

where r is the approximate ratio, $\pi(K)$ is the regret function, K is the number of iterations.

Similar to Lipschitz smoothness [12], one-sided-smoothness can control the approximation ratio and the complexity of related algorithms. The main method to solve the above problem is continuous greedy. The core of continuous greedy is to maximize the multilinear extension function. Submodularity has some beautiful properties for the multilinear continuous extension. For example, monotone concavity along a fixed direction. The property was often used to bound a Taylor expansion in algorithm analysis. Since nonsubmodular multilinear extensions don't have this property, Mehrdad et al proposed a "OSS property" condition to propose an alternative bound based on Taylor series. When the OSS function is monotone, Mehrdad et al. [2] provided a tight $(1 - 1/e^{(1-\beta)(\beta/(\beta+1))^{2\eta}})$ approximation for the maximization problem under downwards-closed convex polytope constraint. For the nonmonotone OSS maximization problem, the research of the algorithm mainly focus on the nonmonotone DR-submodular maximization problem, that is $\eta = 0$. It is difficult to maximize a non-monotone continuous DR-submodular functions. Bian et al and Niazadeh et al.[13, 14] have given methods to maximize non-monotone diminishing return (DR) submodular functions and they got the same 1/2-approximation guarantee. Both algorithms come from the double greedy framworks in [15, 16]. In 2019, Bian et al. [3] provided a 1/e approximation algorithm on downwards-closed convex sets. For general convex sets containing origin, Dürr et al. [4] gave a $\frac{1}{3\sqrt{3}}$ approximation solution.

In order to optimize the more general problems (i.e. $\eta > 0$), we propose a nonmonotone Frank Wolfe method. This method comes from the technique of solving convex optimization problems [17, 18], and has been widely introduced into the research of submodular optimization and machine learning problems [12, 19]. The core of the Frank Wolfe is to approximate the original function by a linear gradient function at each iteration point x_k. If the objective function is monotone, then the gradient function is nonnegative and it was often used to analyze the approximate solution [20]. However, it is not useful for the nonmonotone case. To overcome this problem, we use several optimization tools that include the OSS function is (θ, ϖ)-continuous, gradient is μ-bounded. The above two optimization tools help us establish the connection between two adjacent iteration points and the optimal solution (i.e. $\psi(x^{k+1}) \geq (1-\rho)\psi(x^k) + \rho(1-\rho\mu^2)^{t^k/\rho\mu^2}\psi(z^*) - \frac{L(D\rho\mu)^2}{2}$, The notations used see Table 1). This connection is very helpful for our design of nonmonotone Frank Wolfe algorithm for OSS maximization problems.

Main Contributions: This paper mainly studies whether the approximate ratio $1/e$ can be guaranteed theoretically under the general OSS maximization problem. We design the nonmonotone Jump-Start Frank Wolfe (JSFK) algorithm by applying several techniques, including Jump-Start, nonmonotone Frank-Wolfe, (θ, ϖ)-continuous, μ-bounded gradient. JSFK algorithm iterates by constantly accessing the optimal linear gradient function $d^T\nabla\psi(x)$ ($d \in \mathbb{P}, d \leq \varpi - x$), and finally outputs the solution. We theoretically prove that JSFK algorithm has a $1/e$ approximation ratio with a regret $O(\frac{L(\mu D)^2}{2K})$ for any OSS function maximization problem, and it needs at least $O(K)$ iterations. Because $\mu = \left(\frac{\beta}{\beta+1}\right)^{-2\eta}$, we choose the larger parameter η, the algorithm will receive the more regret (ie, the result of the algorithm will get worse as the parameter increases).

Main differences and limitations: In this paper, our main work is to extend the study of nonmonotone continuous DR-submodular functions under the same approximation ratio guarantee. The analysis idea of the algorithm comes from the Algorithm 2 (in [3]) and Algorithm 1 (in [4]), but there are essential differences and limitations.

- Main differences:
  - The analysis of Algorithm 2 (in [3]) and Algorithm 1 (in [4]) strictly depends on the concaveness in the nonnegative direction of the continuous DR-submodular function.
  - Our algorithm iterates from a non-zero point. (It is possible to set the polyhedron constraint to not contain the origin, unless for some k, $x^k \geq z^*$ hold, where z^* denotes the optimal solution).
- Limitations:
  - The analysis of our algorithm requires stronger gradient conditions: μ-bounded continuous.
  - Compared with Algorithm 2 (in [3]), the Jump-Start Frank Wolfe algorithm need to pay more regrets to obtain the same approximate ratio.
  - If the parameter η is too large, the optimization model and its algorithm that we study are meaningless

1.1 Organization

The remainder of this paper is organized as follows. Section 2 introduced some definitions and necessary lemmas for the later algorithm designs. In Section 3 , a nonmonotone algorithm are proposed for the deterministic nonmonotone OSS problem. The corresponding theoretical analysis for their efficiency are also provided. The last section concludes this work.

2 Preliminaries

The notations used in this paper are lised in Table 1.

In this section, we would offer some notations and definitions, which are used throughout the whole paper. Let ψ be a normalized, nonnegative OSS function. For $\forall u, u' \in [0,1]^n$, $u \leq u'$ if and only if $u_i \leq u'_i$ holds. $u \vee u'$ denotes the coordinate-wise maximum of u and u', and $u \wedge u'$ denotes the coordinate-wise minimum.

Definition 1. Down-closed: *For $\forall x \in \mathbb{P}$, if $\mathbf{0} \leq x' \leq x$, then $x' \in \mathbb{P}$.*

To obtain a constant approximation ratio in polynomial time, the constraint set needs to have the property of downwards-closed. Therefor the downwards-closed condition is very important for approximation algorithms analysis process.

Lemma 1. *([2]) $\psi : [0,1]^n \to \mathbb{R}_+$ is OSS on $[x, x+\epsilon v]$, then we have*

$$v^T \nabla \psi(x+\epsilon v) \leq \left(\frac{\|x+\epsilon v\|_1}{\|x\|_1}\right)^{2\eta} v^T \nabla \psi(x). \quad (3)$$

where $x, v \in [0,1]^n, x \neq \mathbf{0}$ and $\epsilon > 0$ such that $(x+\epsilon v) \in [0,1]^n$.

Table 1. Symbol Description

β, η, ϵ	: Some parameters greater than zero are given in advance;
μ	: $\mu = \left(\frac{\beta}{\beta+1}\right)^{-2\eta}$
$\mathbb{P}$	: Polyhedron: such as linear function polyhedron $Ax \leq b$, matrix polyhedron, etc.;
ϖ	: The upper bound vector of $x \in P$;
u	: A vector belongs to $[0,1]^n$;
D	: The diameter of P, where $D := \max_{x,x' \in \mathbb{P}} \|x - x'\|$, and $D \leq \|\varpi\|$;
L	: The Lipschitz parameter.

Definition 2. *([21])* **Lipschitz smooth:** *A gradient function $\psi : [0,1]^n \to \mathbb{R}$ is Lipschitz smooth if for all $x, y \in [0,1]^n$, it holds that*

$$\|\nabla\psi(x) - \nabla\psi(x')\| \leq L\|x - x'\|. \tag{4}$$

Lemma 2. *A gradient function $\psi : [0,1]^n \to \mathbb{R}$ has L-Lipschitz gradients if for all $x, x' \in [0,1]^n$, then we have*

$$|\psi(x') - \psi(x) - \langle\nabla\psi(x), x' - x\rangle| \leq \frac{L}{2}\|x' - x\|^2. \tag{5}$$

We defer to the proofs to the full version.

In order to obtain the convergence approximation solution, we need the gradient function to be Lipschitz continuous. In addition, in order to overcome the difficulties caused by the nonmonotone property of the objective function, the objective function also needs to satisfy the following two conditions.

Definition 3. (θ, ϖ)**-continuous:** *Given $\theta \in [\mathbf{0}, \varpi]$ (where ϖ is the upper bound vector of $x \in \mathbb{P}$). Then a function $\psi : [0,1]^n \to \mathbb{R}_+$ is (θ, ϖ)-continuous if for any $x \in [\mathbf{0}, \theta]$, $y \in \mathbb{P}$, the following inequality*

$$\psi(x \vee y) \geq \left(1 - \left[\min_{i\in[n]} \frac{\varpi_i}{\theta_i}\right]^{-1}\right)\psi(y). \tag{6}$$

holds.

Definition 4. μ**-bounded continuous:** *A gradient function $\nabla\psi(x) : [0,1]^n \to \mathbb{R}$ is μ-bounded continuous if for any $x, y, y' \in \mathbb{P}$, the following inequality holds*

$$|\langle\nabla\psi(x), y\rangle| \leq \mu \cdot |\langle\nabla\psi(x), y'\rangle| \tag{7}$$

where $\mu = \left(\frac{\beta}{\beta+1}\right)^{-2\eta}$.

The gradient function with a μ-bound provides a bound for the variation of the gradient function to avoid the function being too singular.

Definition 5. *([22]) Continuous DR-submodular functions: A continuously twice differentiable function $\psi : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_+$ is DR-submodular if it satisfies*

$$\psi(ke_i + x) - \psi(x) \geq \psi((k+l)e_i + x) - \psi(le_i + x),$$

where $k, l \in \mathbb{R}_+$ and $x, (ke_i + x), ((k+l)e_i + x) \in \mathbb{R}^n_{\geq 0}$.

Submodularity has some beautiful properties for the multilinear continuous extension. For example, monotone concavity along a fixed direction. The property was often used to bound a Taylor expansion in algorithm analysis. Next we need consider some questions: no monotonicity? no submodularity? no unidirectional concave? Mehrdad et al. proposed a "OSS-property" condition which guarantees an alternative bound based on Taylor series. But they did not consider the case where the function is nonmonotone.

3 Jump-Start Frank Wolfe Algorithm for nonmonotone Setting

For the nonmonotone OSS problems, we propose Jump-Start Frank Wolfe algorithm, our algorithm mainly uses the Frank Wolfe skill in convex optimization. That is, the following linear optimization problem is solved in each iteration of the algorithm

$$\max_{d \in \mathbb{P}} d^T \nabla \psi(x_k).$$

The algorithm can be briefly described in three parts: (i) Choose an initial feasible solution $x^0 = \beta \arg\max_{x \in P, x \leq \varpi} \|x\|_1, \beta \to 0, \beta \in [0,1)$; (ii) Identify the Iteration direction by $Frank\ Wolfe$ skill; (iii) Travel an acceptable distance in the selected direction.

In Algorithm 1, t^k is the cumulative step size. When $t^k = 1$, x^K is the convex combination of feasible solutions. x^K must be a feasible solution. ϖ is used to bound the growth of x^k (For any $k < K$, we can get $x^k \leq \varpi[1 - (1 - \rho\mu^2)^{t^k/\rho\mu^2}]$), where $\mu = \left(\frac{\beta}{\beta+1}\right)^{-2\eta}$. The following is a specific analysis.

Algorithm 1 JSFW: Jump-Start Frank Wolfe Algorithm($\psi, \mathbb{P}, \beta, \eta, K, \varpi$)

Input: η-OSS function $\psi : [0,1]^n \cap \mathbb{P} \to \mathbb{R}_+$; $\mathbb{P}$: down-closed convex polytope; ϖ: the given bounded vector (i.e. for any $x \in \mathbb{P}$, $x \leq \varpi$ holds).
Parameter: $\beta \in (0,1]$, $\eta \geq 0$, step size $\rho = 1/K$, $K \approx O(n^2)$. Let $\mu = (\beta/\beta+1)^{-2\eta}$.
Output: x^K.

1: $t^0 \to \beta$.
2: $k = 0$.
3: $x^0 = \beta \arg\max_{x \in \mathbb{P}, x \leq \varpi} \|x\|_1$.
4: **while** $t^k < 1$, **do**
5: $\quad d^k = \arg\max_{d \in \mathbb{P}, d \leq \varpi - x^k} d^T \nabla \psi(x^k)$, $\rho_k = \min\{\rho, 1 - t^k\}$,
6: $\quad x^{k+1} = x^k + \mu^2 \rho_k d^k$, $t^{k+1} = t^k + \mu^2 \rho_k$,
7: $\quad k \leftarrow k + 1$.
8: **end while**

Lemma 3. *Assume* $x^0 = \beta \arg\max_{x \in P, x \leq \varpi} \|x\|_1$. *For any* $k < K$, *it holds,*

$$x^k \leq \varpi \left[1 - \left(1 - \rho\mu^2\right)^{t^k/\rho\mu^2}\right], \tag{8}$$

where $\mu = \left(\frac{\beta}{\beta+1}\right)^{-2\eta}$.

We defer to the proofs to the full version.

Lemma 4. *Let* z^* *denote the optimal solution, then for all* x^k *in Algorithm 1, we have*

$$\psi(x^k \vee z^*) - \psi(x^k) \leq \mu|(x^k \vee z^* - x^k)^T \nabla\psi(x^k)|. \tag{9}$$

We defer to the proofs to the full version.

Lemma 5. *For any* $k < K$, *the following inequality holds*

$$\psi(x^{k+1}) \geq (1-\rho)\psi(x^k) + \rho(1-\rho\mu^2)^{t^k/\rho\mu^2}\psi(z^*) - \frac{L(D\rho\mu)^2}{2}, \tag{10}$$

where $\mu = \left(\frac{\beta}{\beta+1}\right)^{-2\eta}$, z^* *denotes the optimal solution.*

Proof. Let $l^k := x^k \vee z^* - x^k$, then the following two conditions hold: 1) $l^k \leq \varpi - x^k$; 2) $l^k \in \mathbb{P}$, since property of downwards-closed. So l^k is a feasible solution in Algorithm 1). Because $\nabla\psi(x)$ is Lipschitz continuous, from Lemma 2, we have

$$\begin{aligned}
\psi(x^{k+1}) - \psi(x^k) &\geq \rho\mu^2\langle\nabla\psi(x^k), d^k\rangle - \frac{L(\rho\mu^2)^2}{2}\|d^k\|^2 \\
(D: The\ diameter\ of\ \mathbb{P}) & \\
&\geq \rho\mu^2\langle\nabla\psi(x^k), d^k\rangle - \frac{L(\rho\mu^2)^2}{2}D^2 \\
(d^k\ is\ an\ ascending\ direction) & \\
&\geq \rho\mu^2|\langle\nabla\psi(x^k), d^k\rangle| - \frac{L(\rho\mu^2)^2}{2}D^2 \\
(By\ the\ \mu-bounded) & \\
&\geq \rho\mu|\langle\nabla\psi(x^k), l^k\rangle| - \frac{L(\rho\mu)^2}{2}D^2 \\
&= \rho\mu|\langle\nabla\psi(x^k), x^k \vee z^* - x^k\rangle| - \frac{L(\rho\mu)^2}{2}D^2 \\
(By\ the\ Lemma\ 4) & \\
&\geq \rho\mu[\psi(x^k \vee z^*) - \psi(x^k)]\cdot\mu^{-1} - \frac{L(\rho\mu)^2}{2}D^2 \\
(By\ the\ (\theta,\varpi)-continuity\ and & \\
\lambda = \min_{i\in[n]}\frac{\varpi_i}{\theta_i}) & \\
&\geq \rho[\left(1 - \frac{1}{\lambda}\right)\psi(z^*) - \psi(x^k)] - \frac{L(\rho\mu)^2}{2}D^2 \\
\left(\theta := \varpi\left[1 - \left(1-\rho\mu^2\right)^{t^k/\rho\mu^2}\right]\right) & \\
&= \rho[(1-\rho\mu)^{t^k/\rho\mu^2}\psi(z^*) - \psi(x^k)] - \frac{L(\rho\mu^2)^2}{2}D^2.
\end{aligned}$$

Hence

$$\psi(x^{k+1}) \geq (1-\rho)\psi(x^k) + \rho(1-\rho\mu^2)^{t^k/\rho\mu^2}\psi(z^*) - \frac{L(D\rho\mu)^2}{2}.$$

□

Theorem 1. *Consider Algorithm 1 with uniform step size ρ. For $k = 1, ..., K$ it holds that*

$$\psi(x^{k+1}) \geq t^{k+1}e^{-t^{k+1}}\psi(z^*) - \frac{(k+1)L}{2}(\rho\mu D)^2 - O(\rho^2)\psi(z^*). \tag{11}$$

The larger the parameter η we choose, the more regrets we will receive, because of $\mu = \left(\frac{\beta}{\beta+1}\right)^{-2\eta}$ ($\beta \in (0,1]$).

Proof. Firstly, it holds when $k = 0$ (notice that $t^0 = \beta \to 0$). Assume that it holds for k. Then for $k+1$, from $0 \leq \rho\mu \leq t \leq 1$ and Lemma 5, we get

$$\begin{aligned}
\psi(x^{k+1}) &\geq (1-\rho)\psi(x^k) + \rho(1-\rho\mu^2)^{t^k/\rho\mu^2}\psi(z^*) - \frac{LD^2}{2}\rho^2\mu^2 \\
&\quad (Because\ e^{-t} - O(\rho\mu^2) \leq (1-\rho\mu^2)^{t/\rho\mu^2}) \\
&\geq (1-\rho)\psi(x^k) + \rho\left[e^{-t^k} - O(\rho\mu^2)\right]\psi(z^*) - \frac{LD^2}{2}\rho^2\mu^2 \\
&\geq (1-\rho)\left[t^k e^{-t^k}\psi(z^*) - \frac{kL}{2}(\rho\mu D)^2 - O(\rho^2)\psi(z^*)\right] \\
&\quad +\rho\left[e^{-t^k} - O(\rho\mu^2)\right]\psi(z^*) - \frac{LD^2}{2}\rho^2\mu^2 \\
&= \left[(1-\rho)t^k e^{-t^k} + \rho e^{-t^k}\right]\psi(z^*) - \frac{(\rho\mu D)^2 L}{2}[(1-\rho)k+1] \\
&\quad -[(1-\rho)O(\rho^2) + \rho O(\rho\mu^2)]\psi(z^*) \\
&\geq \left[(1-\rho)t^k e^{-t^k} + \rho e^{-t^k}\right]\psi(z^*) - \frac{(k+1)(\rho\mu D)^2 L}{2} \\
&\quad -O(\rho^2)\psi(z^*).
\end{aligned}$$

Next, let $g(t) = te^{-t}$, the function is is monotonically increasing in $[0,1]$ and $g(t+\rho) - g(t) \leq \rho g'(t)$. Then we get

$$\begin{aligned}
&\left[(1-\rho)t^k e^{-t^k} + \rho e^{-t^k}\right]\psi(z^*) \\
&\geq (t^k+\rho)e^{-(t^k+\rho)}\psi(z^*) \\
&\geq (t^k+\rho\mu^2)e^{-(t^k+\rho\mu^2)}\psi(z^*) \\
&= (t^{k+1})e^{-(t^{k+1})}\psi(z^*).
\end{aligned}$$

So the Theorem 1 holds, that is

$$\psi(x^{k+1}) \geq t^{k+1}e^{-t^{k+1}}\psi(z^*) - \frac{(k+1)L}{2}(\rho\mu D)^2 - O(\rho^2)\psi(z^*)$$

The algorithm termination condition is $t = 1$, and we need about $O(Ln^2)$ number of iterations, when the algorithm terminates, we can get the following solution

$$\psi(x^{out}) \geq e^{-1}OPT - \frac{L(\mu D)^2}{2K} - O(\rho^2)\psi(z^*).$$

Because $\mu = \left(\frac{\beta}{\beta+1}\right)^{-2\eta}$, we choose the larger parameter η, the algorithm will receive the more regret (ie, the result of the algorithm will get worse as the parameter increases). Most of OSS problems are continuous, and our algorithms can be applied directly. However, if the optimization problems are discrete, then we need to round the corresponding non-discrete results. Generally, the traditional rounding techniques mentioned in the paper [2] are all available, the rounding solution will have a loss.

4 Conclusion

In this paper, we design a nonmonotone Jump Start Frank Wolfe algorithm for nonmonotone OSS maximization problem with a down-closed convex polytope constraint. Our algorithm obtains a e^{-1}-approximation solution with a $\frac{L(\mu D)^2}{2K}$ (where $\mu = \left(\frac{\beta}{\beta+1}\right)^{-2\eta}$) regret. If $\eta = 0$, the result is same as the nonmonotone DR-submodular maximization problem in [5] and the result may not be tight. In order to get the convergent approximation solution, the algorithm requires at least $O(K)$ iterations. We choose the larger parameter η, the algorithm will receive more regret. Our paper also provides a good tool to maximize the multilinear extension of some nonmonotone set functions.

Acknowledgements

The first author is supported by National Natural Science Foundation of China (No. 12131003) and General Research Projects of Beijing Educations Committee in China under Grant (No. KM201910005013).

References

1. G. Mehrdad, S. Richard, S. Bruce, Beyond submodular maximization, arXiv preprint arXiv:1904.09216 (2019) 1–60.
2. G. Mehrdad, S. Richard, S. Bruce, Beyond submodular maximization via one-sided smoothness, in: Proceedings of SODA, 2021, pp. 1006–1025.
3. A. Bian, K. Levy, A. Krause, J. Buhmann, Non-monotone continuous dr-submodular maximization: structure and algorithms, in: Proceedings of NeurIPS, 2018, pp. 487–497.
4. C. Dürr, N. Thăng, A. Srivastav, L. Tible, Non-monotone dr-submodular maximization over general convex sets, in: Proceedings of IJCAI, 2021, pp. 2148–2154.
5. C. Chandra, K. Quanrud, Submodular function maximization in parallel via the multilinear relaxation, in: Proceedings of the SODA, 2019, pp. 303–322.
6. F. Radlinski, S. Dumais, Improving personalized web search using result diversification, in: Proceedings of ACM, SIGIR, 2006, pp. 691–692.

7. M. Ghadiri, M. Schmidt, Distributed maximization of submodular plus diversity functions for multilabel feature selection on huge datasets, in: Proceedings of AISTATS, 2019, pp. 2077–2086.
8. S. Zadeh, M. Ghadiri, V. Mirrokni, M. Zadimoghaddam, Scalable feature selection via distributed diversity maximization, in: Proceedings of AAAI, 2017, pp. 2876–2883.
9. Z. Abbassi, V. Mirrokni, M. Thakur, Diversity maximization under matroid constraints, in: Proceedings of ACM SIGKDD, 2013, pp. 32–40.
10. D. Xin, H. Cheng, X. Yan, J. Han, Extracting redundancy-aware top-k patterns, in: Proceedings of ACM SIGKDD, 2006, pp. 444–453.
11. J. Carbonell, J. Goldstein, The use of mmr, diversity-based reranking for reordering documents and producing summaries, in: Proceedings of ACM SIGIR, 1998, pp. 335–336.
12. A. Bian, K. Levy, A. Krause, J. Buhmann, Continuous dr-submodular maximization: structure and algorithms, in: Proceedings of NeurIPS, 2017, pp. 486–496.
13. Y. Bian, J. Buhmann, A. Krause, Optimal continuous dr-submodular maximization and applications to provable mean field inference, in: Proceedings of ICML, 2019, pp. 644–653.
14. R. Niazadeh, T. Roughgarden, J. Wang, Optimal algorithms for continuous non-monotone submodular and dr-submodular amximization, Journal of Machine Learning Research 21 (2020) 1–31.
15. N. Buchbinder, M. Feldman, Deterministic algorithms for submodular maximization problems, ACM Transactions on Algorithms 14 (2018) 1–20.
16. N. Buchbinder, M. Feldman, J. Seffi, R. Schwartz, A tight linear time (1/2)-approximation for unconstrained submodular maximization, SIAM Journal on Computing 44 (2015) 1384–1402.
17. J. Martin, Revisiting frank-wolfe: Projection-free sparse convex optimization, in: Proceedings of ICML, 2013, pp. 427–435.
18. R. Freund, P. Grigas, New analysis and results for the frank-wolfe method, Mathematical Programming 155 (2016) 199–230.
19. M. Zhang, Z. Shen, A. Mokhtari, H. Hassani, A. Karbasi, One sample stochastic frank-wolfe, in: Proceedings of ICAIS, 2020, pp. 4012–4023.
20. M. Feldman, J. Naor, R. Schwartz, A unified continuous greedy algorithm for submodular maximization, in: Proceedings of FOCS, 2011, pp. 570–579.
21. Y. Nesterov, Introductory lectures on convex optimization: a basic course, Vol. 87, Springer Science and Business Media, 2003.
22. Y. Bian, J. Buhmann, A. Krause, Continuous submodular function maximization, arXiv preprint arXiv:2006.13474 (2020) 1–64.