Non-monotone DR-submodular Maximization over General Convex Sets

Christoph Dürr[1] , **Nguyễn Kim Thắng**[2] , **Abhinav Srivastav**[2] and **Léo Tible**[2]

[1]LIP6, Sorbonne University, France
[2]IBISC, Univ Évry, University Paris-Saclay, France
Christoph.Durr@lip6.fr, kimthang.nguyen@univ-evry.fr, abhinavsriva@gmail.com,
leo.tible@ens-cachan.fr

Abstract

Many real-world problems can often be cast as the optimization of DR-submodular functions defined over a convex set. These functions play an important role with applications in many areas of applied mathematics, such as machine learning, computer vision, operation research, communication systems or economics. In addition, they capture a subclass of non-convex optimization that provides both practical and theoretical guarantees.

In this paper, we show that for maximizing non-monotone DR-submodular functions over a *general* convex set (such as up-closed convex sets, conic convex set, etc) the Frank-Wolfe algorithm achieves an approximation guarantee which depends on the convex set. To the best of our knowledge, this is the first approximation guarantee for general convex sets. Finally we benchmark our algorithm on problems arising in machine learning domain with the real-world datasets.

1 Introduction

Submodular (set function) optimization has been widely studied for decades The domain has been under extensive investigation in recent years due to its numerous applications in statistics and machine learning , for example active learning [Golovin and Krause, 2011], viral marketing [Kempe *et al.*, 2003], network monitoring [Gomez Rodriguez *et al.*, 2010], document summarization [Lin and Bilmes, 2011], crowd teaching [Singla *et al.*, 2014], feature selection [Elenberg *et al.*, 2018], deep neural networks [Elenberg *et al.*, 2017], diversity models [Djolonga *et al.*, 2016] and recommender systems [Guillory and Bilmes, 2011]. One of the primary reason for the popularity of submodular functions, is its diminishing returns property. This property has been observed in many settings and is known also as the economy of scale property in economics and operations research.

Continuous diminishing-return submodular (DR-submodular) functions are natural extension of submodular set functions to the continuous setting, specifically by extending the notion of diminishing returns property to continuous domains. These functions have recently gained a significant attention in both, machine learning and theoretical computer science communities [Bach, 2016; Bian *et al.*, 2018; 2017a; Chen *et al.*, 2018] while formulating real-world problems such as non-concave quadratic programming, revenue maximization, log-submodularity and mean inference for Determinantal Point Processes. More examples of this can be found in [Ito and Fujimaki, 2016], [Kulesza *et al.*, 2012], and [Djolonga and Krause, 2014]. Moreover, DR-submodular functions captures a subclass of non-convex functions that enables both exact minimization [Iwata *et al.*, 2001]and approximate maximization [Bian *et al.*, 2017a] in polynomial time.

Previous work on the DR-submodular maximization problem has been studied with crucial assumptions on either additional properties of the functions (e.g., monotonicity) or special structures of the convex sets (e.g., unconstrained or down-closed structures). Though, the majority of real-world problems can be formulated as non-monotone DR-submodular functions over a general convex sets.

In this paper, we investigate the problem of maximizing constrained non-monotone DR-submodular functions. Our result provides an approximation algorithm for maximizing smooth, non-monotone DR-submodular function over general convex sets. As the name suggest, general convex sets includes *conic convex sets, upward-closed convex sets, mixture of covering and packing linear constraints, etc*. Prior to this work, no algorithm with performance guarantee was known for the above mentioned problem.

In this paper, we consider non-negative DR-submodular functions F, that at any point $\boldsymbol{x} \in [0,1]^n$, $F(\boldsymbol{x}) \geq 0$ and the value of $F(\boldsymbol{x})$ and its gradient (denoted by $\nabla F(\boldsymbol{x})$) can be evaluated in the polynomial time.

Given a convex set $\mathcal{K} \subseteq [0,1]^n$, we aim to design an iterative algorithm that produces a solution $\boldsymbol{x}^T$ after T iterations such that

$$F(\boldsymbol{x}^T) \geq \max_{\boldsymbol{x} \in \mathcal{K}} \alpha \cdot F(\boldsymbol{x}) - \beta$$

where α, β are some parameters. Such an algorithm is called an α-*approximation* with *convergence rate* β (that depends on, among others, the number of iterations T).

1.1 Our Contributions

The problem of maximizing a non-monotone DR-submodular function over a general convex set has been proved to be hard.

Specifically, any constant-approximation algorithm for this problem must require super-polynomial many value queries to the function [Vondrák, 2013]. Determining the approximation ratio as a function depending on the problem parameters and characterizing necessary and sufficient regularity conditions that enable efficient approximation algorithm for the problem constitute an important direction.

Exploring the underlying properties of DR-submodularity, we show that the celebrated Frank-Wolf algorithm achieves an approximation ratio of $\left(\frac{1-\min_{\boldsymbol{x}\in\mathcal{K}}\|\boldsymbol{x}\|_\infty}{3\sqrt{3}}\right)$ with the convergence rate of $O\left(\frac{n}{\ln^2 T}\right)$, where T is the number of iterations applied in the algorithm. (The exact convergence rate with parameter dependencies is $O\left(\frac{\beta D^2}{(\ln T)^2}\right)$ where β is the smoothness of function F and D is the diameter of $\mathcal{K}$.) In particular, if the set $\mathcal{K}$ contains the origin $\boldsymbol{0}$ then for arbitrary constant $\epsilon > 0$, after $T = O\left(e^{\sqrt{n/\epsilon}}\right)$ (subexponential) iterations, the algorithm outputs a solution $\boldsymbol{x}^T$ such that $F(\boldsymbol{x}^T) \geq \left(\frac{1}{3\sqrt{3}}\right)\max_{\boldsymbol{x}^*\in\mathcal{K}}F(\boldsymbol{x}^*) - \epsilon$. Note that $\mathcal{K}$ containing the origin $\boldsymbol{0}$ is not necessarily down-closed[1], for example a set $\mathcal{K}$ the intersection of a cone[1] and the hypercube $[0,1]^n$.

As mentioned earlier, due to [Vondrák, 2013], when aiming for approximation algorithms for convex sets beyond the down-closed ones, a super-polynomial number of iterations is unavoidable. To the best of our knowledge, this is the first algorithm approximation guarantee for maximizing non-monotone DR-submodular functions over non-down-closed convex sets.

1.2 Related Work

The problem of submodular (set function) minimization has been studied in [Iwata et al., 2001]. See [Bach and others, 2013] for a survey on connections with applications in machine learning. Submodular maximization is an NP-hard problem. Several approximation algorithms have been given in the offline setting, for example a 1/2-approximation for unconstrained domains [Buchbinder et al., 2015; Buchbinder and Feldman, 2018], a $(1 - 1/e)$-approximation for monotone smooth submodular functions [Calinescu et al., 2011] or a $(1/e)$-approximation for non-motonotone submodular functions on down-closed polytopes [Feldman et al., 2011].

Continuous extension of submodular functions play a crucial role in submodular optimization, especially in submodular maximization, including the multilinear relaxation and the softmax extension. These belong to the class of DR-submodular functions. [Bian et al., 2017b] considered the problem of maximizing monotone DR-functions subject to down-closed convex sets and proved that the greedy method proposed by [Calinescu et al., 2011], which is a variant of the Frank-Wolfe algorithm, guarantees a $(1 - 1/e)$-approximation. It has been observed by [Hassani et al., 2017] that the greedy method is not robust in stochastic settings (where only unbiased estimates of gradients are available). Subsequently, they showed that the gradient meth-

ods achieve $1/2$-approximation in stochastic settings. Maximizing non-monotone DR-submodular functions is harder. Very recently, [Bian et al., 2018] and [Niazadeh et al., 2018] have independently presented algorithms with the same approximation guarantee of 1/2 for the problem of maximizing non-monotone DR-submodular functions over $\mathcal{K} = [0,1]^n$. Both algorithm are inspired by the bi-greedy algorithm in [Buchbinder et al., 2015; Buchbinder and Feldman, 2018]. [Bian et al., 2017a] made a further step by providing an $1/e$-approximation algorithm over down-closed convex sets. Remark again that when aiming for approximation algorithms, the restriction to down-closed polytopes is unavoidable. Specifically, [Vondrák, 2013] proved that any algorithm for the problem over a general base of matroids that guarantees a constant approximation must require exponentially many value queries to the function.

2 Preliminaries and Notations

We introduce some basic notions, concepts and lemmas which will be used throughout the paper. We use boldface letters, e.g., $\boldsymbol{x}, \boldsymbol{z}$ to represent vectors. We denote x_i as the ith entry of $\boldsymbol{x}$ and $\boldsymbol{x}^t$ as the decision vector at time step t. For two n-dimensional vectors $\boldsymbol{x}, \boldsymbol{y}$, we say that $\boldsymbol{x} \leq \boldsymbol{y}$ iff $x_i \leq y_i$ for all $1 \leq i \leq n$. Moreover, $\boldsymbol{x} \vee \boldsymbol{y}$ is defined as a vector such that $(x \vee y)_i = \max\{x_i, y_i\}$ and similarly $\boldsymbol{x} \wedge \boldsymbol{y}$ is a vector such that $(x \wedge y)_i = \min\{x_i, y_i\}$. In the paper, we use the Euclidean norm $\|\cdot\|$ by default (so the dual norm is itself). The infinity norm $\|\cdot\|_\infty$ is defined as $\|\boldsymbol{x}\|_\infty = \max_{i=1}^n |x_i|$.

In the paper, we consider a bounded convex set $\mathcal{K}$ and w.l.o.g. assume that $\mathcal{K} \subseteq [0,1]^n$. We say that $\mathcal{K}$ is *general* if $\mathcal{K}$ is simply a convex set without any particular property. $\mathcal{K}$ is *down-closed* if for any vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ such that $\boldsymbol{y} \in \mathcal{K}$ and $\boldsymbol{x} \leq \boldsymbol{y}$, then $\boldsymbol{x}$ is also in $\mathcal{K}$. A *cone* is a set $\mathcal{C}$ such that if $\boldsymbol{x} \in \mathcal{C}$ then $\alpha\boldsymbol{x} \in \mathcal{C}$ for any constant $\alpha > 0$. Besides, the *diameter* of the convex domain $\mathcal{K}$ (denoted by D) is defined as $\sup_{\boldsymbol{x}, \boldsymbol{y}\in\mathcal{K}}\|\boldsymbol{x} - \boldsymbol{y}\|$.

A function $F : [0,1]^n \to \mathbb{R}_+$ is *DR-submodular* if for all vectors $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ with $\boldsymbol{x} \geq \boldsymbol{y}$, any basis vector $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ and any constant $\alpha > 0$ such that $\boldsymbol{x} + \alpha\boldsymbol{e}_i \in [0,1]^n$, $\boldsymbol{y} + \alpha\boldsymbol{e}_i \in [0,1]^n$, it holds that

$$F(\boldsymbol{x} + \alpha\boldsymbol{e}_i) - F(\boldsymbol{x}) \leq F(\boldsymbol{y} + \alpha\boldsymbol{e}_i) - F(\boldsymbol{y}). \quad (1)$$

This is called *diminishing-return (DR)* property. Note that if function F is differentiable then the diminishing-return (DR) property (1) is equivalent to

$$\nabla F(\boldsymbol{x}) \leq \nabla F(\boldsymbol{y}) \qquad \forall \boldsymbol{x} \geq \boldsymbol{y} \in [0,1]^n. \quad (2)$$

Moreover, if F is twice-differentiable then the DR property is equivalent to all of the entries of its Hessian being nonpositive, i.e., $\frac{\partial^2 F}{\partial x_i \partial x_j} \leq 0$ for all $1 \leq i, j \leq n$. A differentiable function $F : \mathcal{K} \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be β-*smooth* if for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{K}$, we have

$$F(\boldsymbol{y}) \leq F(\boldsymbol{x}) + \langle \nabla F(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{\beta}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2 \quad (3)$$

or equivalently,

$$\left\|\nabla F(\boldsymbol{x}) - \nabla F(\boldsymbol{y})\right\| \leq \beta\|\boldsymbol{x} - \boldsymbol{y}\|. \quad (4)$$

A property of DR-submodular functions, that we will use in our paper, is the following.

---

[1]We refer to Section 2 for a formal definition

Lemma 1 ([Hassani *et al.*, 2017], proof of Theorem 4.2). *For every $x, y \in \mathcal{K}$ and any DR-submodular function $F : [0,1]^n \to \mathbb{R}_+$, it holds that*

$$\langle \nabla F(x), y - x \rangle \geq F(x \vee y) + F(x \wedge y) - 2F(x).$$

Consequently, as F is non-negative, $\langle \nabla F(x), y - x \rangle \geq F(x \vee y) - 2F(x)$.

3 DR-submodular Maximization over General Convex Sets

In this section, we consider the problem of maximizing a DR-submodular function over a general convex set. Approximation algorithms [Bian *et al.*, 2017b; 2017a] have been presented and all of them are adapted variants of the Frank-Wolfe method. However, those algorithms require that the convex set is down-closed. This structure is crucial in their analyses in order to relate their solution to the optimal solution. Using some property of DR-submodularity (specifically, Lemma 1), we show that beyond the down-closed structure, the Frank-Wolf algorithm guarantees an approximation solution for general convex sets. Below, we present the pseudocode of our variant of the Frank-Wolfe algorithm.

Algorithm 1 Frank-Wolfe Algorithm

1: Let $x^1 \leftarrow \arg\min_{x \in \mathcal{K}} \|x\|_\infty$.
2: **for** $t = 1$ to T **do**
3:     Compute $v^t \leftarrow \arg\max_{v \in \mathcal{K}} \langle \nabla F(x^{t-1}), v \rangle$
4:     Update $x^t \leftarrow (1 - \eta_t)x^{t-1} + \eta_t v^t$ with step-size $\eta_t = \frac{\delta}{t H_T}$ where H_T is the T-th harmonic number and δ represents the constant $(\ln 3)/2$.
5: **end for**
6: **return** x^T

First, we show that during the execution of the algorithm, the following invariant is maintained.

Lemma 2. *It holds that $1 - x_i^t \geq e^{-\delta - O(1/\ln^2 T)} \cdot (1 - x_i^1)$ for every $1 \leq i \leq n$ and every $1 \leq t \leq T$.*

Proof. Fix a dimension $i \in [n]$. For every $t \geq 2$, we have

$$
\begin{aligned}
1 - x_i^t &= 1 - (1 - \eta_t)x_i^{t-1} - \eta_t v_i^t \\
&\quad \text{(using the Update step from Algorithm 1)} \\
&\geq 1 - (1 - \eta_t)x_i^{t-1} - \eta_t \quad\quad (1 \geq v_i^t) \\
&= (1 - \eta_t)(1 - x_i^{t-1}) \\
&\geq e^{-\eta_t - \eta_t^2} \cdot (1 - x_i^{t-1}) \\
&\quad \text{(since } 1 - u \geq e^{-u - u^2} \text{ for } 0 \leq u < 1/2)
\end{aligned}
$$

Using this recursion, we have,

$$
\begin{aligned}
1 - x_i^t &\geq e^{-\sum_{t'=2}^{t} \eta_{t'} - \sum_{t'=2}^{t} \eta_{t'}^2} \cdot (1 - x_i^1) \\
&\geq e^{-\delta - O(1/\ln^2 T)} \cdot (1 - x_i^1)
\end{aligned}
$$

because $\sum_{t'=2}^{t} \eta_{t'}^2 = \sum_{t'=2}^{t} \frac{\delta^2}{t'^2 H_T^2} = O(1/(\ln T)^2)$ (since δ is a constant, $H_T = \Theta(\ln T)$, $\sum_{t'=2}^{t} \frac{1}{t'^2} = O(1)$); and $\sum_{t'=2}^{t} \eta_{t'} = \frac{\delta}{H_T} \sum_{t'=2}^{t} \frac{1}{t'} \leq \delta \frac{H_t}{H_T} \leq \delta$. $\square$

The following lemma was first observed in [Feldman *et al.*, 2011] and was generalized in [Bian *et al.*, 2017a, Lemma 3].

Lemma 3 ([Feldman *et al.*, 2011; Bian *et al.*, 2017a]). *For every $x, y \in \mathcal{K}$ and any DR-submodular function $F : [0,1]^n \to \mathbb{R}^+$, it holds that $F(x \vee y) \geq (1 - \|x\|_\infty) F(y)$.*

Theorem 1. *Let $\mathcal{K} \subseteq [0,1]^n$ be a convex set and let $F : [0,1]^n \to \mathbb{R}_+$ is a non-monotone β-smooth DR-submodular function. Let D be the diameter of $\mathcal{K}$. Then Algorithm 1 yields a solution $x^T \in \mathcal{K}$ such that the following inequality holds:*

$$
\begin{aligned}
F(x^T) &\geq \left(\frac{1}{3\sqrt{3}} \right) \left(1 - \min_{x \in \mathcal{K}} \|x\|_\infty \right) \cdot \max_{x \in \mathcal{K}} F(x) \\
&\quad - O\left(\frac{\beta D^2}{(\ln T)^2} \right).
\end{aligned}
$$

Proof. Let $x^* \in \mathcal{K}$ be the maximum solution of F. Let $r = e^{-\delta - O(1/\ln^2 T)} \cdot (1 - \max_i x_i^1)$. Note that from Lemma 2, it follows that $(1 - \|x^t\|_\infty) \geq r$ for every t. Next we present a recursive formula in terms of $F(x^t)$ and $F(x^*)$:

$$
\begin{aligned}
&2F(x^{t+1}) - rF(x^*) \\
&= 2F\big((1 - \eta_{t+1})x^t + \eta_{t+1}v^{t+1}\big) - rF(x^*) \\
&\quad\quad \text{(Using the Update step from Algorithm 1)}
\end{aligned}
$$

$$
\begin{aligned}
&\geq 2F(x^t) - rF(x^*) \\
&\quad - 2\eta_{t+1}\langle \nabla F\big((1 - \eta_{t+1})x^t + \eta_{t+1}v^{t+1}\big), (x^t - v^{t+1})\rangle \\
&\quad - \beta(\eta_{t+1})^2 \|x^t - v^{t+1}\|^2 \\
&\quad\quad (\beta\text{-smoothness as defined by Inequality (3))}
\end{aligned}
$$

$$
\begin{aligned}
&= 2F(x^t) - rF(x^*) + 2\eta_{t+1}\langle \nabla F(x^t), (v^{t+1} - x^t)\rangle \\
&\quad + 2\eta_{t+1}\langle \nabla F\big((1 - \eta_{t+1})x^t + \eta_{t+1}v^{t+1}\big)(v^{t+1} - x^t)\rangle \\
&\quad - 2\eta_{t+1}\langle \nabla F(x^t), (v^{t+1} - x^t)\rangle \\
&\quad - \beta(\eta_{t+1})^2 \|v^{t+1} - x^t\|^2
\end{aligned}
$$

$$
\begin{aligned}
&\geq 2F(x^t) - rF(x^*) + 2\eta_{t+1}\langle \nabla F(x^t), (v^{t+1} - x^t)\rangle \\
&\quad - 2\eta_{t+1}\big\| \nabla F(a^t) - \nabla F(x^t)\big\|\big\|(v^{t+1} - x^t)\big\| \\
&\quad - \beta(\eta_{t+1})^2 \|v^{t+1} - x^t\|^2 \\
&\quad\quad (\text{where } a^t = (1 - \eta_{t+1})x^t + \eta_{t+1}v^{t+1}) \\
&\quad\quad (\text{Using Cauchy-Schwarz})
\end{aligned}
$$

$$
\begin{aligned}
&\geq 2F(x^t) - rF(x^*) + 2\eta_{t+1}\langle \nabla F(x^t), (v^{t+1} - x^t)\rangle \\
&\quad - 3\beta(\eta_{t+1})^2 \|v^{t+1} - x^t\|^2 \\
&\quad\quad (\beta\text{-smoothness as defined by Inequality (4))}
\end{aligned}
$$

$$
\begin{aligned}
&\geq 2F(x^t) - rF(x^*) + 2\eta_{t+1}\langle \nabla F(x^t), x^* - x^t\rangle \\
&\quad - 3\beta(\eta_{t+1})^2 \|v^{t+1} - x^t\|^2 \quad\quad (\text{definition of } v^{t+1})
\end{aligned}
$$

$$
\geq 2F(x^t) - rF(x^*) + 2\eta_{t+1}\Big(F(x^* \vee x^t) - 2F(x^t)\Big)
$$

$$-3\beta(\eta_{t+1})^2\|\boldsymbol{v}^{t+1} - \boldsymbol{x}^t\|^2 \qquad\qquad\text{(Lemma 1)}$$

$$\geq 2F(\boldsymbol{x}^t) - rF(\boldsymbol{x}^*) + 2\eta_{t+1}\left(rF(\boldsymbol{x}^*) - 2F(\boldsymbol{x}^t)\right)$$
$$-3\beta(\eta_{t+1})^2\|\boldsymbol{v}^{t+1} - \boldsymbol{x}^t\|^2 \qquad\qquad\text{(Lemma 3)}$$

$$\geq (1 - 2\eta_{t+1})\left(2F(\boldsymbol{x}^t) - rF(\boldsymbol{x}^*)\right) - 3\beta(\eta_t)^2 D^2,$$

where D is the diameter of $\mathcal{K}$.

Let $h^t = 2F(\boldsymbol{x}^t) - rF(\boldsymbol{x}^*)$. By the previous inequality and the choice of η_t, we have

$$h^{t+1} \geq (1 - 2\eta_{t+1})h^t - 3\beta(\eta_t)^2 D^2$$
$$= (1 - 2\eta_{t+1})h^t - O\left(\frac{\beta\delta^2 D^2}{t^2(\ln T)^2}\right).$$

where we used the facts that $H_T = \Theta(\ln T)$. Note that the constants hidden by the O-notation above is small. Therefore,

$$h^T \geq \prod_{t=2}^{T}(1 - 2\eta_t)h^1 - O\left(\frac{\beta\delta^2 D^2}{(\ln T)^2}\right)\sum_{t=1}^{T}\frac{1}{t^2}$$
$$\geq e^{-2\sum_{t=2}^{T}\eta_t - 4\sum_{t=2}^{T}\eta_t^2}\cdot h^1 - O\left(\frac{\beta\delta^2 D^2}{t^2(\ln T)^2}\right)$$
$$\text{(since }(1 - u) \geq e^{-u-u^2}\text{ for }0 \leq u < 1/2)$$
$$\geq e^{-2\delta - O(1/\ln^2 T)}\cdot h^1 - O\left(\frac{\beta\delta^2 D^2}{(\ln T)^2}\right)$$

where the last inequality holds because $\sum_{t=2}^{T}\eta_t^2 = \sum_{t=2}^{T}\frac{\delta^2}{t^2 H_T^2} = O(1/(\ln T)^2)$ (as δ is a constant) and $\sum_{t=2}^{T}\eta_t = \frac{\delta}{H_T}\sum_{t=2}^{T}\frac{1}{t} \leq \delta$. Hence,

$$2F(\boldsymbol{x}^T) - rF(\boldsymbol{x}^*)$$
$$\geq e^{-2\delta - O\left(1/(\ln T)^2\right)}\cdot\left(2F(\boldsymbol{x}^1) - rF(\boldsymbol{x}^*)\right)$$
$$- O\left(\frac{\beta\delta^2 D^2}{(\ln T)^2}\right)$$

which implies,

$$F(\boldsymbol{x}^T)$$
$$\geq \frac{r}{2}\left(1 - e^{-2\delta - O\left(1/(\ln T)^2\right)}\right)F(\boldsymbol{x}^*) - O\left(\frac{\beta\delta^2 D^2}{(\ln T)^2}\right)$$

By the definition of r, the multiplicative factor of $F(\boldsymbol{x}^*)$ in the right-hand side is

$$\frac{e^{-\delta - O(1/\ln^2 T)}\cdot\left(1 - e^{-2\delta - O\left(1/(\ln T)^2\right)}\right)}{2}(1 - \max_i x_i^1)$$

Note that for T sufficiently large, $O(1/(\ln T)^2) \ll 1$. Optimizing this factor by choosing $\delta = \frac{\ln 3}{2}$, we get

$$F(\boldsymbol{x}^T) \geq \left(\frac{1}{3\sqrt{3}}\right)(1 - \max_i x_i^1)F(\boldsymbol{x}^*) - O\left(\frac{\beta D^2}{(\ln T)^2}\right)$$

and the theorem follows. $\qquad\qquad\square$

Remark. The above theorem also indicates that the performance of the algorithm, specifically the approximation ratio, depends also on the initial starting point $\boldsymbol{x}^1$. This matches to the observation in non-convex optimization that the quality of the output solution depends on, among others, the initialization. The following corollary is a direct consequence of the theorem for a particular family of convex sets $\mathcal{K}$ beyond the down-closed ones. Note that for any convex set $\mathcal{K} \subseteq [0, 1]^n$, the diameter $D \leq \sqrt{n}$.

Corollary 1. *If $0 \in \mathcal{K}$, then the guarantee in Theorem 1 can be written as:*

$$F(\boldsymbol{x}^T) \geq \left(\frac{1}{3\sqrt{3}}\right)\max_{\boldsymbol{x}^* \in \mathcal{K}}F(\boldsymbol{x}^*) - O\left(\frac{\beta n}{(\ln T)^2}\right).$$

where the starting point $\boldsymbol{x}^1 = 0$.

4 Experiments

In this section, we validate our algorithm for non-monotone DR submodular optimization on both real-world and synthetic datasets. Our experiments are broadly classified into following two categories:

1. We compare our algorithm against two previous known algorithms for maximizing non-monotone DR-submodular function over *down-closed* polytopes mentioned in [Bian *et al.*, 2017a]. Recall that our algorithm applies to a more general setting where the convex set is not required to be down-closed.

2. Next, we show the performance of our algorithm for maximizing non-monotone DR-submodular function over *general* polytopes. Recall that no previous algorithm was known to have performance guarantees for this problem.

All experiments are performed in MATLAB using CPLEX optimization tool on MAC OS version 10.14[2].

4.1 Performance over Down-closed Polytopes

Here, we benchmark the performance of our variant of the Frank-Wolfe algorithm from Section 3 against the previous known two algorithms for maximizing continuous DR submodular functions over *down-closed* polytopes mentioned in [Bian *et al.*, 2017a]. We considered QUADPROGIP, which is a global solver for non-convex quadratic programming, as a baseline. We run all the algorithms for 100 iterations. All the results are the average of 20 repeated experiments. For the sake of completion, we describe below the problem and different settings used. We follow closely the experimental settings from [Bian *et al.*, 2017a], and adapted their source codes to our algorithm.

Quadratic Programming. As a state-of-the-art global solver, we used QUADPROGIP to find the global optimum

---
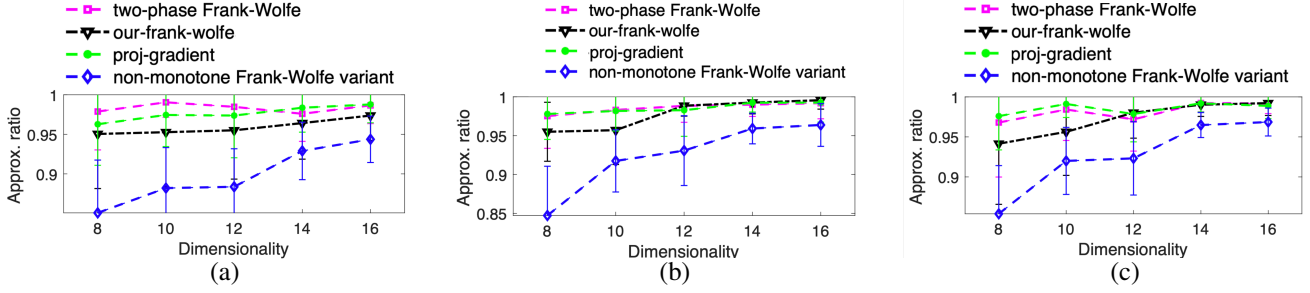
Figure 1: Uniform distribution with down-closed polytope and (a) $m = \lfloor 0.5n \rfloor$ (b) $m = n$ (c) $m = \lfloor 1.5n \rfloor$.

which is used to calculate the approximation ratios. Our problem instances are synthetic quadratic objectives with down-closed polytope constraints, i.e.,

$$f(x) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{H} \boldsymbol{x} + \boldsymbol{h}^\top \boldsymbol{x} + c$$

and $\mathcal{K} = \{\boldsymbol{x} \in \mathbb{R}_+^n | \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \leq \boldsymbol{u}, \boldsymbol{A} \in \mathbb{R}_+^{m \times n}, \boldsymbol{b} \in \mathbb{R}_+^m\}$. Note that in previous sections, we have assumed w.l.o.g that $\mathcal{K} \subseteq [0,1]^n$. By scaling our results hold as well for the general box constraint $\boldsymbol{x} \leq \boldsymbol{u}$, provided the entries of $\boldsymbol{u}$ are upper bounded by a constant.

Both objective and constraints were randomly generated, using the following two ways:

Uniform Distribution. $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ is a symmetric matrix with uniformly distributed entries in $[-1, 0]$; $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has uniformly distributed entries in $[\mu, \mu + 1]$, where $\mu = 0.01$ is a small positive constant in order to make entries of $\boldsymbol{A}$ strictly positive for down-closed polytopes.

Exponential Distribution. Here, the entries of $\boldsymbol{H}$ and $\boldsymbol{A}$ are sampled from exponential distributions $Exp(\lambda)$ where given a random variable $y \geq 0$, the probability density function of y is defined by $\lambda e^{-\lambda y}$, and for $y < 0$, its density is fixed to be 0. Specifically, each entry of $\boldsymbol{H}$ is sampled from $Exp(1)$. Each entry of $\boldsymbol{A}$ is sampled from $Exp(0.25) + \mu$, where $\mu = 0.01$ is a small positive constant.

We set $\boldsymbol{b} = \boldsymbol{1}^m$, and $\boldsymbol{u}$ to be the tightest upper bound of $\mathcal{K}$ by $u_j = \min_{i \in [m]} \frac{b_i}{A_{ij}}, \forall j \in [n]$. In order to make f non-monotone, we set $\boldsymbol{h} = -0.2 * \boldsymbol{H}^\top \boldsymbol{u}$. To make sure that f is non-negative, first we solve the problem $\min_{\boldsymbol{x} \in \mathcal{P}} \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{H} \boldsymbol{x} + \boldsymbol{h}^\top \boldsymbol{x}$ using QUADPROGIP. Let the solution to be $\hat{x}$, then we set $c = -f(\hat{x}) + 0.1 * |f(\hat{x})|$.

The approximation ratios w.r.t. the dimension n are plotted in Figures 1 and 2 for the two distributions. In each figure, we set the number of constraints to be $m = \lfloor 0.5n \rfloor$, $m = n$ and $m = \lfloor 1.5n \rfloor$, respectively.

We can observe that our version of Frank-Wolfe (denoted our-frank-wolfe) has comparable performance with the state-of-the-art algorithms when optimizing DR-submodular functions over down-closed convex sets. Note that the performance is clearly consistent with the proven approximation guarantee of $1/e$ shown in [Bian *et al.*, 2017a]. We also show that the performance of our algorithm is consistent with the proven approximation guarantee of $(1/3\sqrt{3})$ for down-closed convex sets.

4.2 Performance over General Polytopes

We consider the problem of revenue maximization on a (undirected) social network graph $G(V, E)$ where V the set of vertices (users) and E is set of edges (connections between users). Additionally, there are weight w_{ij} of the edge between vertex i and vertex j. The goal is to offer for free or advertise a product to users so that the revenue increases through their "word-of-mouth" effect on others. If one invests x unit of cost on a user $i \in V$, the user i becomes an advocate of the product (independently from other users) with probability $1 - (1 - p)^x$ where $p \in (0, 1)$ is a parameter. Intuitively, it means that for investing a unit cost to i, we have an extra chance that the user i becomes an advocate with probability p [Soma and Yoshida, 2017].

Let $S \subset V$ be a set of users who advocate for the product. Note that S is random set. Then the revenue with respect to S is defined as $\sum_{i \in S} \sum_{j \in V \setminus S} w_{ij}$. Let $f : \mathbb{R}_+^{|V|} \to \mathbb{R}_+$ be the expected revenue obtained in this model, that is

$$f(\boldsymbol{x}) = \mathbb{E}_S \Big[\sum_{i \in S} \sum_{j \in V \setminus S} w_{ij} \Big]$$
$$= \sum_i \sum_{j: i \neq j} w_{ij} (1 - (1 - p)^{x_i})(1 - p)^{x_j}$$

It has been shown that f is a non-monotone DR-submodular function [Soma and Yoshida, 2017]. We imposed a minimum and a maximum investment constraint on the problem such that $0.25 \leq \sum_i x_i \leq 1$. This, in addition with $x_i \geq 0$, forms a general (non-down-closed) convex set $\mathcal{K} := \{\boldsymbol{x} \in \mathbb{R}_+^{|V|} : 0.25 \leq \sum_i x_i \leq 1, x_i \geq 0 \ \forall i\}$. In our experiments, we used the Advogato network with 6.5K users (vertices) and 61K connections (edges). We set $p = 0.0001$.

In Figure 3, we show the performance of the Frank-Wolfe algorithm (as mentioned in Section 3) and commonly used Gradient Ascent algorithm. It is imperative to note that no performance guarantee is known for the Gradient Ascent algorithm for maximizing a non-monotone DR-submodular function over a general constraint polytope. We can clearly observe that the Frank-Wolfe algorithm performs at least as good as the commonly used Gradient Ascent algorithm.

5 Conclusion

In this paper, we have provided performance guarantees for the problems of maximizing non-monotone submodular/DR-
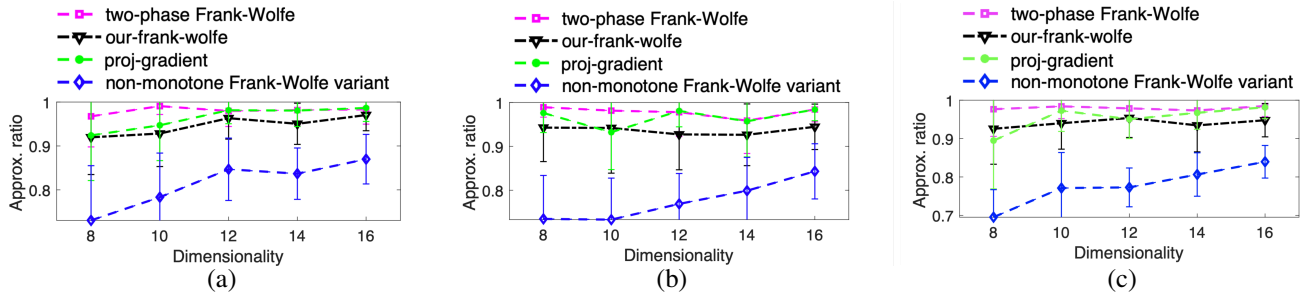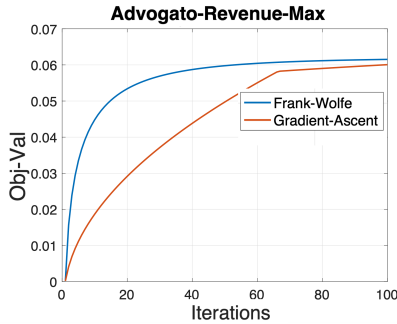
Figure 2: Exponential distribution with down-closed polytope and (a) $m = \lfloor 0.5n \rfloor$, (b) $m = n$, (c) $m = \lfloor 1.5n \rfloor$

Figure 3: Revenue Maximization on Advogato dataset

submodular functions over convex sets. This result is complemented by experiments in different contexts. Moreover, the results give raise to the question of designing online algorithms for non-monotone DR-submodular maximization over a general convex set. Characterizing necessary and sufficient regularity conditions/structures that enable efficient algorithm with approximation and regret guarantees is an interesting direction to pursue.

Acknowledgements

References

[Bach and others, 2013] Francis Bach et al. Learning with submodular functions: A convex optimization perspective. *Foundations and Trends® in Machine Learning*, 6(2-3):145–373, 2013.

[Bach, 2016] Francis Bach. Submodular functions: from discrete to continuous domains. *Mathematical Programming*, pages 1–41, 2016.

[Bian *et al.*, 2017a] Andrew An Bian, Kfir Levy, Andreas Krause, and Joachim M. Buhmann. Continuous DR-submodular maximization: Structure and algorithms. In *NeurIPS*, 2017.

[Bian *et al.*, 2017b] Andrew An Bian, Baharan Mirzasoleiman, Joachim Buhmann, and Andreas Krause. Guaranteed non-convex optimization: Submodular maximiza-

tion over continuous domains. In *AISTATS*, pages 111–120, 2017.

[Bian *et al.*, 2018] An Bian, Joachim M Buhmann, and Andreas Krause. Optimal DR-submodular maximization and applications to provable mean field inference. In *NeurIPS*, 2018.

[Buchbinder and Feldman, 2018] Niv Buchbinder and Moran Feldman. Deterministic algorithms for submodular maximization problems. *ACM TALG*, 14(3):32, 2018.

[Buchbinder *et al.*, 2015] Niv Buchbinder, Moran Feldman, Joseph Seffi, and Roy Schwartz. A tight linear time (1/2)-approximation for unconstrained submodular maximization. *SIAM Computing*, 44(5):1384–1402, 2015.

[Calinescu *et al.*, 2011] Gruia Calinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM J. Computing*, 40(6):1740–1766, 2011.

[Chen *et al.*, 2018] Lin Chen, Hamed Hassani, and Amin Karbasi. Online continuous submodular maximization. In *AISTAT*, 2018.

[Djolonga and Krause, 2014] Josip Djolonga and Andreas Krause. From map to marginals: Variational inference in bayesian submodular models. In *NeurIPS*, pages 244–252, 2014.

[Djolonga *et al.*, 2016] Josip Djolonga, Sebastian Tschiatschek, and Andreas Krause. Variational inference in mixed probabilistic submodular models. In *NeurIPS*, pages 1759–1767, 2016.

[Elenberg *et al.*, 2017] Ethan Elenberg, Alexandros G Dimakis, Moran Feldman, and Amin Karbasi. Streaming weak submodularity: Interpreting neural networks on the fly. In *NeurIPS*, pages 4044–4054, 2017.

[Elenberg *et al.*, 2018] Ethan R Elenberg, Rajiv Khanna, Alexandros G Dimakis, Sahand Negahban, et al. Restricted strong convexity implies weak submodularity. *The Annals of Statistics*, 46(6B):3539–3568, 2018.

[Feldman *et al.*, 2011] Moran Feldman, Joseph Naor, and Roy Schwartz. A unified continuous greedy algorithm for submodular maximization. In *FOCS*, pages 570–579, 2011.

[Golovin and Krause, 2011] Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications

in active learning and stochastic optimization. *J. of AI Research*, 42:427–486, 2011.

[Gomez Rodriguez *et al.*, 2010] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. In *SIGKDD*, pages 1019–1028, 2010.

[Guillory and Bilmes, 2011] Andrew Guillory and Jeff A Bilmes. Simultaneous learning and covering with adversarial noise. In *ICML*, volume 11, pages 369–376, 2011.

[Hassani *et al.*, 2017] Hamed Hassani, Mahdi Soltanolkotabi, and Amin Karbasi. Gradient methods for submodular maximization. In *NeurIPS*, pages 5841–5851, 2017.

[Ito and Fujimaki, 2016] Shinji Ito and Ryohei Fujimaki. Large-scale price optimization via network flow. In *NeurIPS*, pages 3855–3863, 2016.

[Iwata *et al.*, 2001] Satoru Iwata, Lisa Fleischer, and Satoru Fujishige. A combinatorial strongly polynomial algorithm for minimizing submodular functions. *J. ACM*, 48(4):761–777, 2001.

[Kempe *et al.*, 2003] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *SIGKDD*, pages 137–146, 2003.

[Kulesza *et al.*, 2012] Alex Kulesza, Ben Taskar, et al. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(2–3):123–286, 2012.

[Lin and Bilmes, 2011] Hui Lin and Jeff Bilmes. A class of submodular functions for document summarization. In *ACL*, pages 510–520, 2011.

[Niazadeh *et al.*, 2018] Rad Niazadeh, Tim Roughgarden, and Joshua R Wang. Optimal algorithms for continuous non-monotone submodular and dr-submodular maximization. In *NeurIPS*, 2018.

[Singla *et al.*, 2014] Adish Singla, Ilija Bogunovic, Gábor Bartók, Amin Karbasi, and Andreas Krause. Near-optimally teaching the crowd to classify. In *ICML*, pages 154–162, 2014.

[Soma and Yoshida, 2017] Tasuku Soma and Yuichi Yoshida. Non-monotone DR-submodular function maximization. In *AAAI*, pages 898–904, 2017.

[Vondrák, 2013] Jan Vondrák. Symmetry and approximability of submodular maximization problems. *SIAM J. Computing*, 42(1):265–304, 2013.