An improved approximation algorithm for maximizing a DR-submodular function over a convex set

Donglei Du\textsuperscript{1}, Zhicheng Liu\textsuperscript{2}, Chenchen Wu \textsuperscript{*3}, Dachuan Xu\textsuperscript{4}, and Yang Zhou\textsuperscript{5}

\textsuperscript{1}Faculty of Management, University of New Brunswick, Fredericton, New Brunswick, E3B 9Y2, Canada, E-mail: ddu@unb.ca
\textsuperscript{2}College of Taizhou, Nanjing Normal University, Taizhou, 225300, P. R. China, E-mail: manlzhic@163.com
\textsuperscript{3}College of Science, Tianjin University of Technology, No.391, Binshui West Street, Xiqing District, Tianjin, P. R. China, E-mail: wuchenchen_tjut@163.com
\textsuperscript{4}Beijing Institute for Scientific and Engineering Computing, Beijing University of Technology, Beijing 100124, P.R. China, E-mail: xudc@bjut.edu.cn
\textsuperscript{5}School of Mathematics and Statistics, Shandong Normal University, Jinan, Shandong, P. R. China, E-mail: zhyg1212@163.com

March 29, 2022

Abstract

Maximizing a DR-submodular function subject to a general convex set is an NP-hard problem arising from many applications in combinatorial optimization and machine learning. While it is highly desirable to design efficient approximation algorithms under this general setting where neither the objective function is monotonic nor the feasible set is down-closed, unfortunately, Vondrák (2013) shows that such a problem admits no constant approximation under the value oracle model. Our main contribution is to present a \(\frac{1}{4} = 0.25\)-approximation Frank-Wolfe type of algorithm with a sub-exponential time-complexity under the value oracle model, improving a previous approximation ratio of \(\frac{2}{3\sqrt{3}} \approx 0.1924\) with the same order of time-complexity by Durr et al. (2021).

Keyword: DR-submodular, \(L\)-smooth, approximation algorithm

1 Introduction

The main problem we are interested in this work is as follows:

\[ x^* \in \arg \max_{x \in P \subseteq [0,1]^n} F(x). \]

In the above problem, we make the following assumptions:

(i) \( P \subseteq [0,1]^n \) is a convex set, \( 0 = (0, \ldots, 0) \in P \) and \( F(0) = 0 \).

*Corresponding author
(ii) $F : P \mapsto \mathbb{R}^+$ is $L$-smooth: $\forall x, y \in P,$
\[
|F(y) - F(x)| - \langle \nabla F(x), y - x \rangle \leq \frac{L}{2} \|y - x\|^2 \iff \|\nabla F(x) - \nabla F(y)\| \leq L \|x - y\|.
\]

\[
F(x) + \langle \nabla F(x), y - x \rangle - \frac{L}{2} \|y - x\|^2 \leq F(y) \leq F(x) + \langle \nabla F(x), y - x \rangle + \frac{L}{2} \|y - x\|^2
\]

(iii) $F : [0,1]^n \mapsto \mathbb{R}^+$ is a non-negative DR-submodular function: $\forall x \geq y \in [0,1]^n; \forall a > 0 :$ $x + ae_i, y + ae_i \in [0,1]^n$,
\[
F(x + ae_i) - F(x) \geq F(y + ae_i) - F(y), \forall i = 1, \ldots, n
\]
\[
\nabla F(x) \geq \nabla F(y).
\]

(iv) Three oracles are available: (i) computing the function value $F(x)$ at any given point $x \in P$, (ii) computing its gradient $\nabla F(x)$ at any given point $x \in P$, and (iii) maximizing a linear function over $P$. \hfill \Box

Note that previous research studying similar problems have been assuming either monotonicity of the objective function $F$ and/or down-closeness of the feasible set $P$. The main reasons are two-fold: on the negative side, lack of either assumption usually means no constant approximation algorithm exists under the value oracle model, as shown by Vondrák (2013); and on the positive side, constant approximation algorithms normally exist when one of the two assumptions holds for submodular maximization problems under many different constraints (see e.g., Buchbinder and Feldman (2018) for an excellent survey on this line of research among many others.).

In this work, we consider the problem without any of the two assumptions. This problem was first studied in (Durr et al., 2021) who design an $\alpha$-approximation algorithm with convergence rate $\beta$:
\[
F(x) \geq \alpha F(x^*) - \beta.
\]

In particular, Durr et al. (2021) show that, under the previous assumptions (i)-(iv), there exists a Frank-Wolfe type of algorithm with
\[
\alpha = \frac{1}{3\sqrt{3}} \approx 0.1924, \beta = O\left(\frac{nL}{\ln^2 T}\right).
\]

where $T$ is the number iterative steps in the algorithm. Therefore, after $T = O\left(e^{\sqrt{\frac{nL}{\epsilon}}}\right)$ iterations, we can obtain $\beta = \epsilon$. Note that the sub-exponential convergence rate (namely, $T = 2^{o(n)}$) is necessary due to the aforementioned inapproximability result in (Vondrák, 2013).

This work contributes along this line of research to improve the approximation ratio to $1/4$ while maintaining the same order of convergence rate:
\[
\alpha = \frac{1}{4} = 0.25, \beta = O\left(\frac{nL}{\ln^2 T}\right). \tag{1}
\]
Therefore, after \( T = O \left( e^{\sqrt{\frac{\sqrt{\varepsilon}}{n}}} \right) \) iterations, our algorithm makes the \( \beta \)-term arbitrarily small: \( \beta = \varepsilon \) with an improved approximation ratio of 1/4.

This problem finds applications in price optimization (Ito and Fujimaki, 2016) and revenue maximization on a (undirected) social network graph (a.k.a., optimal seeding) (Soma and Yoshida, 2017). Interested readers are refereed to (Niazadeh et al., 2018; Bian et al., 2020) and references therein for many applications of this type of submodular maximization problems.

The rest of this paper is organized as follows. In Section 2, we present the Frank-Wolfe algorithm and analyze its approximation ratio along with its convergence rate. Concluding remarks are provided in Section 3.

## 2 The Frank-Wolfe algorithm

Given a DR-submodular function \( F : [0, 1]^n \rightarrow \mathbb{R} \), and a convex set \( P \subseteq [0, 1]^n \), the Frank-Wolfe algorithm is summarized as follows:

**Algorithm 1 Frank-Wolfe algorithm**

1: procedure \( \text{FW-Dis}(F, P, T) \)
2: \[ x(0) = 0 = (0, \ldots, 0) \in \mathbb{R}^n \] \hspace{0.5cm} \( \triangleright \) Initialization: \( 0 \in P \) due to assumption
3: \hspace{0.5cm} for \( j = 0, 1 \ldots, T - 1 \) do
4: \hspace{1cm} \[ t_j = \frac{1}{1 - \ln \left( 1 + \frac{H_j}{H_T} \right)} - 1 \] \hspace{0.5cm} (2)
5: \hspace{1cm} \[ v(x(t_j)) \in \arg \max_{v \in P} \langle \nabla F (x(t_j)), v \rangle \] \hspace{0.5cm} (3)
6: \hspace{1cm} \[ x(t_{j+1}) = e^{-\frac{1}{1+t_j} - \frac{1}{1+t_{j+1}}} x(t_j) + \left( 1 - e^{\frac{1}{1+t_j} + \frac{1}{1+t_{j+1}}} \right) v(x(t_j)) \] \hspace{0.5cm} (4)
7: \hspace{0.5cm} end for
8: return \( x \in \arg \max_{x(t_j)} F(x(t_j)) \) \hspace{0.5cm} \( \triangleright \) \( t_T = \frac{\ln 2}{1 - \ln 2} \approx 2.258891 \)
9: end procedure

The following comments are in order:

(1) To save space in the future analysis, denote
\[
\sqrt{a_{t_j}} := e^{\frac{t_j}{1+t_j}}.
\]

(2) Note that
\[
(2) \iff \sqrt{a_{t_j}} := e^{\frac{t_j}{1+t_j}} = 1 + \frac{H_j}{H_T}, \tag{5}
\]

where \( H_j \) is the harmonic number:
\[
H_T = \sum_{j=1}^{T} \frac{1}{j} = \Theta (\ln T), \tag{6}
\]
\[
H_{2,T} = \sum_{j=1}^{T} \frac{1}{j^2} = \Theta (1). \tag{7}
\]
We will also use the last quantity $H_{2T}$, which is the finite sum of the reciprocals of the first $T$ squared natural numbers, in the convergence rate analysis (see the proof of Theorem 1 in Section 2.1.3).

(3) Moreover, the step-size in (4) and the notation $a_t$ are related as follows:

$$e^{-\frac{1}{1+t_j} + \frac{1}{1+t_{j+1}}} = \sqrt{\frac{a_t}{a_{t+1}}} \in [0, 1]$$

(4) $T$ is the total number of iterations.

(5) $t_j$ in (2) is an increasing function of $j$ and indicates the time value of $t$ at the start of the $j$-th iteration, satisfying the boundary conditions

$$t_0 = 0, t_T = \frac{1}{1-\ln 2} - 1 = \frac{\ln 2}{1-\ln 2} \approx 2.258891.$$

(6) The solution at each step is feasible as it is the convex combination of feasible solutions $x(t_j), v(t_j) \in P$.

2.1 Analysis

2.1.1 Two lemmas

We first bound $x_i(t_j)$ from above.

**Lemma 1.** The algorithm maintains the following inequality:

$$1 - x_i(t_j) \geq e^{-\frac{t_j}{1+t_j}} = \frac{1}{\sqrt{a(t_j)}}, i = 1, \ldots, n; j = 0, 1, \ldots, T.$$

**Proof.** From (4) in the algorithm, we have

$$1 - x_i(t_{j+1}) = 1 - e^{-\frac{1}{1+t_j} + \frac{1}{1+t_{j+1}}} x_i(t_j) - \left(1 - e^{-\frac{1}{1+t_j} + \frac{1}{1+t_{j+1}}} \right) v(x(t_j))$$

$$\geq 1 - e^{-\frac{1}{1+t_j} + \frac{1}{1+t_{j+1}}} x(t_j) - 1 + e^{-\frac{t_j+1}{1+t_{j+1}} + \frac{1}{1+t_j}}$$

$$= -e^{-\frac{1}{1+t_j} + \frac{1}{1+t_{j+1}}} x(t_j) + e^{\frac{t_j+1}{1+t_{j+1}} - \frac{1}{1+t_j}}$$

$$= e^{-\frac{t_j}{1+t_j} + \frac{1}{1+t_{j+1}}} (1 - x(t_j))$$

$$\geq e^{\sum_{j=0}^{T} \left[-\frac{t_j}{1+t_j} + \frac{1}{1+t_{j+1}}\right]} (1 - x(t_0))$$

$$= e^{-\frac{t_{j+1}}{1+t_{j+1}}} = e^{-\frac{t_j}{1+t_j} + \frac{1}{1+t_{j+1}}} = \frac{1}{\sqrt{a(t_{j+1})}}$$

$\square$

We recall the following known result (Feldman et al., 2011, Lemma 3.5), which is a consequence of the DR-submodularity of $F$. For any DR-submodular function $F : P \mapsto \mathbb{R}^+$, and $\forall x, x^* \in P$, where $P \subseteq [0, 1]^n$, we have

$$F(x \vee x^*) \geq (1 - \|x\|_\infty) F(x^*) = \left(1 - \max_{i=1}^n x_i(t)\right) F(x^*)$$

(9)
With (8)-(9), we are able to bound $F(x(t) \vee x^*)$ from below, which is summarized in the following lemma.

**Lemma 2.**

\[
F(x(t_j) \vee x^*) \geq e^{-\frac{t_j}{1+\sqrt{t_j}}} F(x^*) = \frac{1}{\sqrt{a_{t_j}}} F(x^*)
\]  

(10)

\[
F(x(t_j) \vee x^*) \geq \left(1 - \max_{i=1}^{n} x_i(t_j)\right) F(x^*)
\]

(9)

\[
F(x(t_j) \vee x^*) \geq e^{-\frac{t_j}{1+\sqrt{t_j}}} F(x^*) = \frac{1}{\sqrt{a_{t_j}}} F(x^*)
\]

\[
E(t_j) = a_{t_j} F(x(t_j)) - \sqrt{a_{t_j}} F(x^*), j = 0, 1, \ldots, T,
\]

(11)

2.1.2 Bounding the increment of the Lyapunov function from below

We introduce the following Lyapunov function analyze the approximation ratio and the convergence rate of Algorithm FW-Dis:

\[
\|
\begin{align*}
E(t_{j+1}) - E(t_j) & = a_{t_{j+1}} F(x(t_{j+1})) - a_{t_j} F(x(t_j)) - \left[\sqrt{a_{t_{j+1}}} - \sqrt{a_{t_j}}\right] F(x^*) \\
& = \left[\sqrt{a_{t_{j+1}}} - \sqrt{a_{t_j}}\right] F(x^*) \\
& = a_{t_{j+1}} \left[\left(1 - \frac{a_{t_j}}{a_{t_{j+1}}}\right) \langle \nabla F(x(t_j)), v(x(t_j) - x(t_j)) \rangle - \frac{L}{2} \left(1 - \sqrt{\frac{a_{t_j}}{a_{t_{j+1}}}}\right)^2 \|v(x(t_j) - x(t_j))\|^2\right] \\
& + [a_{t_{j+1}} - a_{t_j}] F(x(t_j)) - \left[\sqrt{a_{t_{j+1}}} - \sqrt{a_{t_j}}\right] F(x^*) \\
& \geq \left[1 - \frac{a_{t_j}}{a_{t_{j+1}}}\right] \langle \nabla F(x(t_j)), v(x(t_j) - x(t_j)) \rangle - \frac{L}{2} \left(1 - \sqrt{\frac{a_{t_j}}{a_{t_{j+1}}}}\right)^2 \|v(x(t_j) - x(t_j))\|^2 \\
& + [a_{t_{j+1}} - a_{t_j}] F(x(t_j)) - \left[\sqrt{a_{t_{j+1}}} - \sqrt{a_{t_j}}\right] F(x^*) \\
& \geq \left[1 - \frac{a_{t_j}}{a_{t_{j+1}}}\right] \langle \nabla F(x(t_j)), v(x(t_j) - x(t_j)) \rangle - \frac{L}{2} \left(1 - \sqrt{\frac{a_{t_j}}{a_{t_{j+1}}}}\right)^2 \\
& + [a_{t_{j+1}} - a_{t_j}] F(x(t_j)) - \left[\sqrt{a_{t_{j+1}}} - \sqrt{a_{t_j}}\right] F(x^*)
\end{align*}
\]

Proof. We can always assume that $E(t_j) \leq 0$ is true for $\forall j \in [T-1]$, otherwise we can notice that $F(x(t_j)) \geq \frac{1}{\sqrt{a_{t_j}}} F(x^*) > \frac{1}{4} F(x^*)$ for some $j \in [T-1]$ so that the approximation guarantee is already achieved.
\begin{align*}
(3) \quad a_{t+1} & \left[ 1 - \sqrt{\frac{a_t}{a_{t+1}}} \right] \langle \nabla F(x(t)), x^* - x(t) \rangle - \frac{nL}{2} \left( 1 - \frac{a_t}{a_{t+1}} \right)^2 \\
+ & \left[ a_{t+1} - a_t \right] F(x(t)) - \left[ \sqrt{a_{t+1}} - \sqrt{a_t} \right] F(x^*) \\
(12) \quad a_{t+1} & \left[ 1 - \frac{a_t}{a_{t+1}} \right] \left( 1 - \sqrt{\frac{a_t}{a_{t+1}}} \right) F(x^*) - \frac{nL}{2} \left( 1 - \frac{a_t}{a_{t+1}} \right)^2 \\
+ & \left[ a_{t+1} - a_t \right] F(x(t)) - \left[ \sqrt{a_{t+1}} - \sqrt{a_t} \right] F(x^*) \\
= & \sqrt{a_t} \left[ 1 - \frac{a_t}{a_{t+1}} \right] \frac{a_{t+1}}{a_t} - \frac{a_{t+1}}{a_t} + 1 \right] F(x^*) - a_t \left[ 2 \sqrt{a_{t+1}} + 2 \sqrt{a_t} - 1 \right] F(x(t)) \\
- & \frac{nL}{2} \left( \sqrt{a_{t+1}} - \sqrt{a_t} \right)^2 \\
= & \sqrt{a_t} \left[ \frac{a_{t+1}}{a_t} - 1 \right]^2 F(x^*) - \sqrt{a_t} F(x(t)) \right) - \frac{nL}{2} \left( \sqrt{a_{t+1}} - \sqrt{a_t} \right)^2 \\
E(x(t)) \leq 0 \quad \geq - \frac{nL}{2} \left( \sqrt{a_{t+1}} - \sqrt{a_t} \right)^2
\end{align*}

2.1.3 Approximation ratio and convergence rate

**Theorem 1.** Algorithm FW-DiS outputs a solution \(x_T\) satisfying

\[
F(x(t_T)) \geq \frac{1}{4} F(x^*) - O \left( \frac{nL}{\ln^2 T} \right)
\]

after \(T\) iterations. Consequently,

\[
F(x(t_T)) \geq \frac{1}{4} F(x^*) - \varepsilon
\]

after \(T = O \left( e^{\sqrt{\frac{nL}{\varepsilon}}} \right)\) iterations.

**Proof.** When Algorithm FW-DiS terminates, we have

\[
t_T = \frac{1}{1 - \ln 2} - 1 \approx 2.258891,
\]
and hence
\[ a_0 = 1, \sqrt{a_{t_j}} = e^{t_j/t_T} = e^{(t_j - 1)/(t_T - t_j)} = 2. \]
Therefore summing up inequality (13) over \( j = 0, 1, \ldots, T \), we obtain
\[
E(t_T) - E(t_0) \overset{F(0)=0}{=} a_{t_j}F(x(t_T)) - (\sqrt{a_{t_j}} - 1)F(x^*) = 
4F(x(t_T)) - F(x^*) 
\geq \frac{nL}{2} \sum_{j=0}^{T-1} (\sqrt{a_{t_{j+1}}} - \sqrt{a_{t_j}})^2
\]
\[
= -\frac{nL}{2} \sum_{j=0}^{T-1} \left(1 + \frac{H_{j+1}}{H_T} - 1 \right)^2
= -\frac{nL}{2} \sum_{j=1}^{T} \frac{1}{H_T^2} \sum_{j=1}^{T} \frac{1}{j^2} = -\frac{nL}{2} \frac{H_{2T}}{H_T^2}
\]
\[
\overset{(5)}{=} -O \left(\frac{nL}{\ln^2 T} \right).
\]
So we have
\[
F(x(t_T)) \geq \frac{1}{4} F(x^*) - O \left(\frac{nL}{\ln^2 T} \right).
\]
Consequently,
\[
F(x(t_T)) \geq \frac{1}{4} F(x^*) - \varepsilon,
\]
after \( T = O \left( e^{\sqrt{\frac{nlT}{2}}} \right) \) iterations.

As another corollary, if \( F \) is the multilinear extension of a submodular function, then Feldman and Naor (2013, Lemma 2.3.7.) shows that \( F \) is \( L \)-smooth with
\[
L = O \left(n^2 \right) F(x^*)
\]
Plugging this \( L \) into the above, we have
\[
F(x(t_T)) \geq \left(\frac{1}{4} - O \left(\frac{n^3}{\ln^2 T} \right) \right) F(x^*).
\]
Consequently,
\[
F(x(t_T)) \geq \left(\frac{1}{4} - \varepsilon \right) F(x^*)
\]
after \( T = O \left( e^{\sqrt{\frac{nlT}{2}}} \right) \) iterations.

3 Conclusions

The assumptions of \( 0 \in P \) and \( F(0) = 0 \) can both be relaxed. Instead of initializing both algorithms with the starting point \( x(0) = 0 \), we use \( x(0) = \arg \min_{x \in P} \|x\|_\infty \). Then almost the same arguments can be carried to yield the following starting-point-dependent approximation ratio. We have
\[
F(x(t_T)) \geq \frac{1}{4} F(0) \geq \frac{1}{4} \left(1 - \min_{x \in P \|x\|_\infty} \right) F(x^*) - O \left(\frac{nL}{\ln^2 T} \right).
\]
An open question is whether the approximation \( 1/4 \) in (1) is the best possible with a sub-exponential convergent rate?
Acknowledgements

The first author’s research is supported by the NSERC grant (No. 283106), and NSFC grants (Nos. 11771386 and 11728104). The third author’s research is supported by the NSFC grants (No. 11971349). The fourth author’s research is supported by the NSFC grants (No. 12131003) and Beijing Natural Science Foundation Project (No. Z200002). The fifth author’s research is supported by the NSFC grants (No. 12001335).

References

Bian, Y., Buhmann, J. M., and Krause, A. (2020). Continuous submodular function maximization. arXiv preprint arXiv:2006.13474.

Buchbinder, N. and Feldman, M. (2018). Submodular functions maximization problems. In Handbook of Approximation Algorithms and Metaheuristics, Second Edition, pages 753–788. Chapman and Hall/CRC.

Durr, C., Thang, N., Srivastav, A., and Tible, L. (2021). Non-monotone dr-submodular maximization over general convex sets. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, pages 2148–2154.

Feldman, M., Naor, J., and Schwartz, R. (2011). A unified continuous greedy algorithm for submodular maximization. In 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science, pages 570–579. IEEE.

Feldman, M. and Naor, S. (2013). Maximization problems with submodular objective functions. PhD thesis, Computer Science Department, Technion.

Hassani, H., Soltanolkotabi, M., and Karbasi, A. (2017). Gradient methods for submodular maximization. arXiv preprint arXiv:1708.03949.

Ito, S. and Fujimaki, R. (2016). Large-scale price optimization via network flow. Advances in Neural Information Processing Systems, 29.

Niazadeh, R., Roughgarden, T., and Wang, J. (2018). Optimal algorithms for continuous non-monotone submodular and dr-submodular maximization. In Advances in Neural Information Processing Systems, pages 9594–9604.

Soma, T. and Yoshida, Y. (2017). Non-monotone dr-submodular function maximization. In Thirty-First AAAI conference on artificial intelligence.

Vondrák, J. (2013). Symmetry and approximability of submodular maximization problems. SIAM Journal on Computing, 42(1):265–304.