Guaranteed Non-convex Optimization: Submodular Maximization over Continuous Domains

Andrew A. Bian, Baharan Mirzasoleiman, Joachim M. Buhmann, Andreas Krause
Department of Computer Science, ETH Zurich
{ybian, baharanm, jbuhmann}@inf.ethz.ch, krausea@ethz.ch

Abstract

Submodular continuous functions are a category of (generally) non-convex/non-concave functions with a wide spectrum of applications. We characterize these functions and demonstrate that they can be maximized efficiently with approximation guarantees. Specifically, I) we propose the weak DR property that gives a unified characterization of the submodularity of all set, lattice and continuous functions; II) for maximizing monotone DR-submodular continuous functions subject to down-closed convex constraints, we propose a Frank-Wolfe style algorithm with \((1 - 1/e)\)-approximation, and sub-linear convergence rate; III) for maximizing general non-monotone submodular continuous functions subject to box constraints, we propose a DoubleGreedy algorithm with \(1/3\)-approximation.

Submodular continuous functions naturally find applications in various real-world settings, including influence and revenue maximization with continuous assignments, sensor energy management, multi-resolution data summarization, facility location, etc. Experiments show that the proposed algorithms efficiently generate superior solutions compared to baseline algorithms.

1 Introduction

Non-convex optimization delineates the new frontier in machine learning, arising in numerous learning tasks from training deep neural networks to latent variable models. Understanding, which classes of objectives can be tractably optimized, remains a central challenge. In this paper, we investigate a class of generally non-convex/non-concave functions—submodular continuous functions, and derive algorithms for approximately optimizing them with strong approximation guarantees.

Submodularity is a structural property usually associated with set functions, with important implications for optimization. Optimizing submodular set functions has found numerous applications in machine learning \([12, 10, 4, 2, 5]\). Submodular set functions can be efficiently minimized \([9]\), and there are strong guarantees for approximate maximization \([13, 11]\). Even though submodularity is most widely considered in the discrete realm, the notion can be generalized to arbitrary lattices \([7]\). Recently, \([1]\) showed how results from submodular set function minimization can be lifted to the continuous domain. In this paper, we further pursue this line of investigation, and demonstrate that results from submodular set function maximization can be generalized as well. Note that the underlying concepts associated with submodular function minimization and maximization are quite distinct, and both require different algorithmic treatment and analysis techniques.

As motivation for our inquiry, we illustrate how submodular continuous maximization captures various applications, ranging from influence and revenue maximization, to sensor energy management, and non-convex/non-concave quadratic programming. The details are deferred to Appendix A. We then present two guaranteed algorithms: The first, based on the Frank-Wolfe \([6]\) and continuous greedy \([18]\) algorithm, applies to monotone DR-submodular functions, and provides a \((1 - 1/e)\) \(\hat{\epsilon}\).
Table 1: Comparison of properties of convex and submodular continuous functions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Convex function $g(\cdot)$, $\lambda \in [0, 1]$</th>
<th>Submodular continuous function $f(\cdot)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0^{th}$ order</td>
<td>$\lambda g(x) + (1 - \lambda) g(y) \geq g(\lambda x + (1 - \lambda) y)$</td>
<td>$f(x) + f(y) \geq f(x \vee y) + f(x \wedge y)$</td>
</tr>
<tr>
<td>$1^{st}$ order</td>
<td>$g(y) - g(x) \geq (g(y), x - x)$</td>
<td>\text{weak DR (this work, Definition 2.1)}</td>
</tr>
<tr>
<td>$2^{nd}$ order</td>
<td>$\nabla^2 g(x) \succeq 0$ (positive semi-definite)</td>
<td>$\frac{\partial^2 f}{\partial x_i \partial x_j} \leq 0, \forall i \neq j$</td>
</tr>
</tbody>
</table>

For a background of submodular optimization and related work please see Appendix B. We use $E = \{e_1, e_2, \ldots, e_n\}$ as the ground set, $\chi_i$ as the characteristic vector for element $e_i$. We use $x \in \mathbb{R}^E$ and $x \in \mathbb{R}^n$ interchangeably to indicate a $n$-dimensional vector, $x(i)$ means the $i$-th element of $x$, and $x|x(i)\rightarrow k$ means setting the $i$-th element of $x$ to be $k$ while keeping all others unchanged.
Algorithm 1: Frank-Wolfe for monotone DR-submodular function maximization

Input: \( \max_{x \in \mathcal{P}} f(x), \mathcal{P} \) is down-closed convex set in the positive orthant, prespecified stepsize \( \gamma \in (0, 1] \)

1. \( x^0 \leftarrow 0, t \leftarrow 0, k \leftarrow 0; \) \hfill /\kappa: \text{iteration index}
2. \textbf{while} \( t < 1 \) \textbf{do}
3. \quad find \( v^k \) s.t. \( \langle v^k, \nabla f(x^k) \rangle \geq \alpha \max_{v \in \mathcal{P}} \langle v, \nabla f(x^k) \rangle - \frac{1}{2} \delta L; \) \hfill \alpha \in (0, 1] \text{ is the multiplicative error level, } \delta \in [0, \bar{\delta}] \text{ is the additive error level}
4. \quad find stepsize \( \gamma_k, \text{ e.g., } \gamma_k \leftarrow \gamma \) or by line search \( \gamma_k \leftarrow \arg \max_{\gamma \in [0,1]} f(x^k + \gamma v^k) \), and set \( \gamma_k \leftarrow \min \{ \gamma_k, 1-t \} \);
5. \quad \( x^{k+1} \leftarrow x^k + \gamma_k v^k, t \leftarrow t + \gamma_k, k \leftarrow k + 1; \)
6. \textbf{Return} \( x^K \); \hfill /\text{assuming there are } K \text{ iterations in total}

or equivalently (if twice differentiable) \( \frac{\partial^2 f}{\partial x(i)^2} \leq 0, \forall i \in E. \)

Lemma 2.2 shows that a twice differentiable function \( f(\cdot) \) is DR-submodular iff \( \forall x \in X, \frac{\partial^2 f}{\partial x(i) \partial x(j)} \leq 0, \forall i, j \in E, \) which in general does not imply concavity of \( f(\cdot). \)

3 Maximizing monotone DR-submodular continuous functions

We present an algorithm for maximizing a monotone DR-submodular continuous function \( f(x) \) subject to a general down-closed convex constraint, i.e., \( \max_{x \in \mathcal{P}} f(x) \). A down-closed convex set \( (\mathcal{P}, \bar{\mathcal{P}}) \) is the convex set \( \mathcal{P} \) associated with a lower bound \( \bar{\mathcal{P}} \in \mathcal{P} \), such that 1) \( \forall y \in \mathcal{P}, \bar{\mathcal{P}} \leq y; \) and 2) \( \forall x \in \mathcal{P}, x \in \mathbb{R}^n, \bar{\mathcal{P}} \leq x \leq y \) implies \( x \in \mathcal{P} \). W.l.o.g., we assume \( \mathcal{P} \) lies in the positive orthant and has the lower-bound \( 0. \) This problem setting captures various real-world applications, e.g., the influence maximization with continuous assignments, sensor energy management, etc. Specifically, for influence maximization, the constraint is a down-closed polytope in the positive orthant \( \mathcal{P} = \{ x \mid 0 \leq x \leq \bar{u}, Ax \leq b, \bar{u} \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \} \). First, the problem is \( NP \)-hard.

Proposition 3.1. The problem of maximizing a monotone DR-submodular continuous function subject to general down-closed polytope constraints is \( NP \)-hard. The optimal approximation ratio is \( (1 - 1/e) \) (up to low-order terms), unless \( \mathcal{P} = \mathcal{P} \).

We summarize the Frank-Wolfe style method in Alg. 1. In each iteration the algorithm uses the linearization of the objective function as a surrogate, and move towards a maximizer of this surrogate objective. The maximizer, i.e., \( v^k = \arg \max_{v \in \mathcal{P}} \langle v, \nabla f(x^k) \rangle \), is used as the update direction in iteration \( k \). Finding such a direction requires maximizing a linear objective at each iteration. We find a proper stepsize \( \gamma_k \) in some way, for example, one can simply set it to be the prespecified stepsize \( \gamma \), or using line search. Then the algorithm update the solution using the stepsize \( \gamma_k \) and go to the next iteration. Note that the Frank-Wolfe algorithm can tolerate both multiplicative error \( \alpha \) and additive error \( \delta \) when solving the linear subproblem (Step 3 of Alg. 1). Setting \( \alpha = 1 \) and \( \delta = 0 \), we recover the error-free case.

DR-submodular functions are non-convex/non-concave in general. However, there is certain connection between DR-submodularity and concavity.

Proposition 3.2. A DR-submodular continuous function \( f(\cdot) \) is concave along any non-negative direction, and any non-positive direction.

Proposition 3.2 implies that the univariate auxiliary function \( g_{x,v}(\xi) := f(x + \xi v), \xi \in \mathbb{R}_+, v \in \mathbb{R}^E_+ \) is concave. As a result, the Frank-Wolfe algorithm can follow a concave direction at each step, which is the main reason it can provide the approximation guarantee. To derive the guarantee, we need assumptions on the non-linearity of \( f(\cdot) \) over the domain \( \mathcal{P} \), which closely corresponds to a Lipschitz assumption on the derivative of \( g_{x,v}(\cdot) \) with parameter \( L > 0 \) in \( [0,1] \).

\[ -\frac{L}{2} \xi^2 \leq g_{x,v}(\xi) - g_{x,v}(0) - \xi \nabla g_{x,v}(0) = f(x + \xi v) - f(x) - \langle \xi v, \nabla f(x) \rangle, \forall \xi \in [0,1] \]
with this category of generally non-convex/non-concave objectives. We deferred further details to Appendix.

Experiments

Theorem 4.2. Assuming the optimal solution to be OPT, the output of Alg. 2 has function value no less than \( \frac{1}{2} f(OPT) - \frac{L}{2}\delta\), where \(\delta\in[0, \delta]\) is the additive error level for each 1-D subproblem.

Experiments

We compared the performance of the proposed algorithms with four baseline methods, on both monotone and non-monotone problem instances, they are: monotone DR-submodular NQP, optimal budget allocation, non-monotone submodular NQP and revenue maximization. The results verified that the Frank-Wolfe and DoubleGreedy methods have strong approximation guarantees and generate superior solutions compared to the baseline algorithms. We defer further details to Appendix F.

Conclusion

We characterized submodular continuous functions, and proposed two approximation algorithms to efficiently maximize them. This work demonstrates that the submodularity structure can ensure guaranteed optimization in the continuous setting, thus allowing to model problems with this category of generally non-convex/non-concave objectives.
References


