Online Continuous Submodular Maximization: From Full-Information to Bandit Feedback

Mingrui Zhang† Lin Chen‡ Hamed Hassani♯ Amin Karbasi‡,♮

† Department of Statistics and Data Science, Yale University, ‡ Department of Electrical Engineering, Yale University
♯ Department of Electrical and Systems Engineering, University of Pennsylvania, ♮ Department of Computer Science, Yale University

Yale

Bandit Submodular Maximization

Submodular Functions

Submodular function: $f: 2^\Omega \to \mathbb{R}_{\geq 0}$ defined on a finite ground set Ω such that for every $A \subseteq B \subseteq \Omega$ and $x \in \Omega \setminus B$, we have

$$f(x|A) \geq f(x|B),$$

where $f(x|A) \triangleq f(A \cup \{x\}) - f(A)$ is a discrete derivative.

Continuous DR-submodular functions are the continuous analogue: Let $F: \mathcal{X} \to \mathbb{R}_{\geq 0}$ be a differentiable function defined on a box $\mathcal{X} \triangleq \prod_{i=1}^d \mathcal{X}_i$, where each $\mathcal{X}_i$ is a closed interval of $\mathbb{R}_{\geq 0}$. We say that F is continuous DR-submodular if for every $x, y \in \mathcal{X}$ that satisfy $x \leq y$ and every $i \in [d] \triangleq \{1, \ldots, d\}$, we have

$$\frac{\partial F}{\partial x_i}(x) \geq \frac{\partial F}{\partial x_i}(y).$$

Bandit Optimization

Online optimization is a repeated two-player game. At each round t:

- The learner chooses an action x_t from a convex set $\mathcal{K} \subseteq \mathbb{R}^n$;
- The adversary chooses a reward function F_t from $\mathcal{F}$, a family of real-valued functions;
- The learner receives a reward $F_t(x_t)$, and observes feedback.

The aim is to maximize the **regret**, *i.e.*, the gap between her accumulated reward and the reward of the best single choice in hindsight

$$\mathcal{R}_T \triangleq \max_{x \in \mathcal{K}} \left\{ \sum_{t=1}^T F_t(x) - \sum_{t=1}^T F_t(x_t) \right\}$$

In the bandit setting, the feedback is only a single real number $F_t(x_t)$.

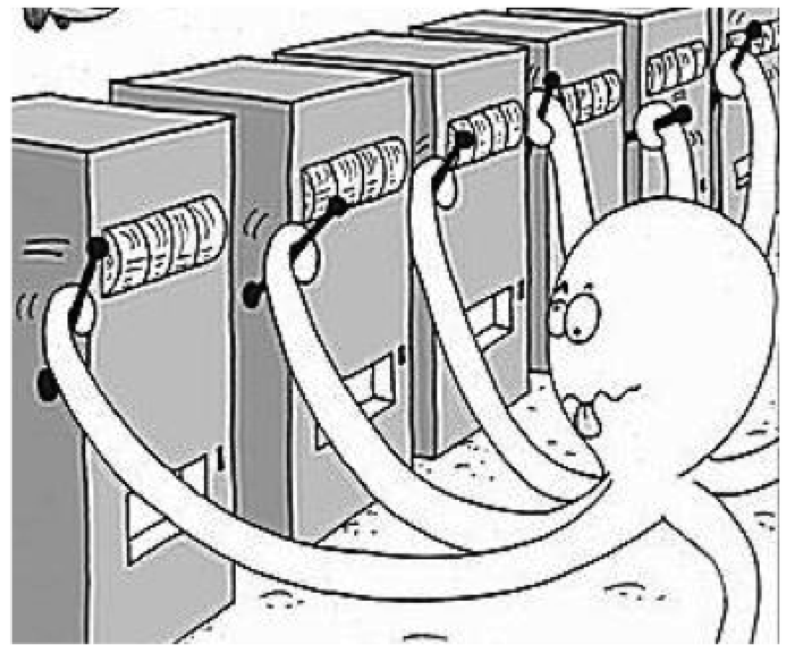


Figure 1: Bandit Problems

However, even in the offline scenario, continuous DR-submodular maximization problem cannot be approximated within a factor of $(1 - 1/e + \epsilon)$ for any $\epsilon > 0$ in polynomial time, unless $RP = NP$ [1]. Therefore, we consider the $(1 - 1/e)$-regret

$$\mathcal{R}_{1-1/e,T} \triangleq (1 - 1/e) \max_{x \in \mathcal{K}} \left\{ \sum_{t=1}^T F_t(x) - \sum_{t=1}^T F_t(x_t) \right\}.$$

Our Contribution

In this paper, we study the following three problems:

- `OCSM`: the Online Continuous DR-Submodular Maximization problem,
- `BCSM`: the Bandit Continuous DR-Submodular Maximization problem, and
- `RBSM`: the Responsive Bandit Submodular Maximization problem.

Table 1: Comparison of previous and our proposed algorithms.

Setting	Algorithm	Stochastic gradient	# of grad. evaluations	$(1-1/e)$-regret
`OCSM`	`Meta-FW`[3]	No	$T^{1/2}$	$O(\sqrt{T})$
	`VR-FW`[2]	Yes	$T^{3/2}$	$O(\sqrt{T})$
	`Mono-FW` (this work)	Yes	1	$O(T^{4/5})$
`BCSM`	`Bandit-FW` (this work)	-	-	$O(T^{8/9})$
`RBSM`	`Responsive-FW` (this work)	-	-	$O(T^{8/9})$

One-shot Online Continuous DR-Submodular Maximization

Key Techniques

Offline Frank-Wolfe (FW) method for maximizing monotone continuous DR-submodular functions: At k-th iteration, solves a linear optimization problem

$$v^{(k)} \leftarrow \arg\max_{v \in \mathcal{K}} \langle v, \nabla F(x^{(k)}) \rangle$$

which is used to update $x^{(k+1)} \leftarrow x^{(k)} + \eta_k v^{(k)}$, where η_k is the step size.

Obstacles of extending FW to online setting:

- To obtain $v_t^{(k)}$, we have to know the gradient in advance.
  Solution: Use K no-regret online linear maximization oracles $\{\mathcal{E}^{(k)}\}, k \in [K]$, with $\langle \cdot, d_t^{(k)} \rangle$ being the objective function, and $v_t^{(k)}$ being the output, where $d_t^{(k)}$ is an estimation of $\nabla F_t(x_t(k))$. By the no-regret property of $\mathcal{E}^{(k)}$, $v_t^{(k)}$ can approximately maximize $\langle \cdot, \nabla F_t(x_t^{(k)}) \rangle$ (First proposed in [2, 3], where $K = T^{3/2}$ stochastic gradients are required for each function).
- Reduce the number of stochastic gradients to 1.
  Idea1: *the blocking procedure.* Divide the objective functions $F_1, \ldots, F_T$ into Q equisized blocks, and define the average function in the q-th block as $\bar{F}_q \triangleq \frac{1}{K} \sum_{k=1}^K F_{(q-1)K+k}$.
  Idea2: *the permutation methods.* Permute the indices $\{(q-1)K+1, \ldots, qK\}$, then we can obtain stochastic gradients of $\bar{F}_q$ at K points (Lines 4 and 5 in `Mono-Frank-Wolfe`).
  Solution (combine Idea1 and Idea2): View the average functions $\bar{F}_1, \ldots, \bar{F}_Q$ as *virtual* objective functions.

Algorithm 1: `Mono-Frank-Wolfe`

Input: constraint set $\mathcal{K}$, horizon T, block size K, online linear maximization oracles on $\mathcal{K}$: $\mathcal{E}^{(1)}, \cdots, \mathcal{E}^{(K)}$, step sizes $\rho_k \in (0,1), \eta_k \in (0,1)$, number of blocks $Q = T/K$

Output: $y_1, y_2, \ldots$

1: **for** $q = 1, 2, \ldots, Q$ **do**
2: $d_q^{(0)} \leftarrow 0$, $x_q^{(1)} \leftarrow 0$
3: For $k = 1, 2, \ldots, K$, let $v_q^{(k)} \in \mathcal{K}$ be the output of $\mathcal{E}^{(k)}$ in round q, $x_q^{(k+1)} \leftarrow x_q^{(k)} + \eta_k v_q^{(k)}$. Set $x_q \leftarrow x_q^{(K+1)}$
4: Let $(t_{q,1}, \ldots, t_{q,K})$ be a random permutation of $\{(q-1)K+1, \ldots, qK\}$
5: For $t = (q-1)K+1, \ldots, qK$, play $y_t = x_q$ and obtain the reward $F_t(y_t)$; find the corresponding $k' \in [K]$ such that $t = t_{q,k'}$, observe $\tilde{\nabla} F_t(x_q^{(k')})$, *i.e.*, $\tilde{\nabla} F_{t_{q,k'}}(x_q^{(k')})$
6: For $k = 1, 2, \ldots, K$, $d_q^{(k)} \leftarrow (1-\rho_k) d_q^{(k-1)} + \rho_k \tilde{\nabla} F_{t_{q,k}}(x_q^{(k)})$, compute $\langle v_q^{(k)}, d_q^{(k)} \rangle$ as reward for $\mathcal{E}^{(k)}$, and feed back $d_q^{(k)}$ to $\mathcal{E}^{(k)}$
7: **end for**

Bandit Continuous DR-Submodular Maximization

Key Techniques

One-point Gradient Estimator: define the δ-smoothed version of a function F as $\hat{F}_\delta(x) \triangleq \mathbb{E}_{v \sim B^d}[F(x + \delta v)]$. Then we have

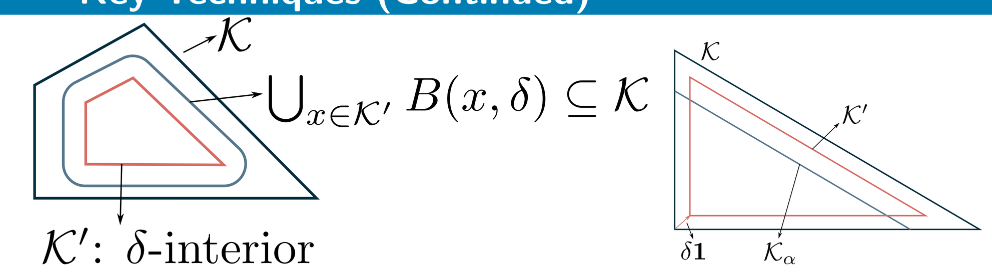
$$\nabla \hat{F}_\delta(x) = \mathbb{E}_{u \sim S^{d-1}} \left[\frac{d}{\delta} F(x + \delta u) u \right].$$

Obstacles of extending `Mono-Frank-Wolfe` to bandit setting:

- The point $x + \delta u$ may fall outside of $\mathcal{K}$.
  Solution: Introduce the notion of δ-*interior*. A set $\mathcal{K}'$ is a δ-interior of $\mathcal{K}$ if it is a *subset* of
  $$\mathrm{int}_\delta(\mathcal{K}) = \{x \in \mathcal{K} | \inf_{s \in \partial \mathcal{K}} d(x, s) \geq \delta\}.$$
  Also define the discrepancy between $\mathcal{K}$ and $\mathcal{K}'$ by
  $$d(\mathcal{K}, \mathcal{K}') = \sup_{x \in \mathcal{K}} d(x, \mathcal{K}'),$$
  Can use the one-point gradient estimator on $\mathcal{K}'$. When F_t is Lipschitz and $d(\mathcal{K}, \mathcal{K}')$ is small, can approximate the optimal total reward on $\mathcal{K}$ ($\max_{x \in \mathcal{K}} \sum_{t=1}^T F_t(x)$) by that on $\mathcal{K}'$ ($\max_{x \in \mathcal{K}'} \sum_{t=1}^T F_t(x)$).
- Lemma 1: Under some regularization assumptions, the set $\mathcal{K}' = (1-\alpha)\mathcal{K} + \delta \mathbf{1}$ is a δ-interior of $\mathcal{K}$ with $d(\mathcal{K}, \mathcal{K}') \leq [\sqrt{d}(\frac{R}{r} + 1) + \frac{R}{r}]\delta$.

Bandit Continuous DR-Submodular Maximization (Continued)

Key Techniques (Continued)



(a) Example of δ-interior (b) Construction of δ-interior

Figure 2: δ-interior

- The one-point estimator requires that the point at which we estimate the gradient (*i.e.*, x) must be identical to the point that we play (*i.e.*, $x + \delta u$), if we ignore the random δu. In `Mono-Frank-Wolfe`, however, we play point x_q but obtain estimated gradient at other points $x_q^{(k')}$ (Line 5).
  Solution: *A biphasic approach* that categorizes the plays into the exploration and exploitation phases.

Algorithm 2: `Bandit-Frank-Wolfe`

Input: smooth radius δ, δ-interior $\mathcal{K}'$ with lower bound $\underline{u}$, horizon T, block size L, exploration steps K, online linear maximization oracles on $\mathcal{K}'$: $\mathcal{E}^{(1)}, \cdots, \mathcal{E}^{(K)}$, step sizes $\rho_k \in (0,1), \eta_k \in (0,1)$, number of blocks $Q = T/L$

Output: $y_1, y_2, \ldots$

1: **for** $q = 1, 2, \ldots, Q$ **do**
2: $d_q^{(0)} \leftarrow 0$, $x_q^{(1)} \leftarrow \underline{u}$
3: For $k = 1, 2, \ldots, K$, let $v_q^{(k)} \in \mathcal{K}'$ be the output of $\mathcal{E}^{(k)}$ in round q, $x_q^{(k+1)} \leftarrow x_q^{(k)} + \eta_k (v_q^{(k)} - \underline{u})$. Set $x_q \leftarrow x_q^{(K+1)}$
4: Let $(t_{q,1}, \ldots, t_{q,L})$ be a random permutation of $\{(q-1)L+1, \cdots, qL\}$
5: **for** $t = (q-1)L+1, \cdots, qL$ **do**
6: If $t \in \{t_{q,1}, \cdots, t_{q,K}\}$, find the corresponding $k' \in [K]$ such that $t = t_{q,k'}$, play $y_t = y_{t_{q,k'}} = x_q^{(k')} + \delta u_{q,k'}$ for F_t (*i.e.*, $F_{t_{q,k'}}$), where $u_{q,k'} \sim S^{d-1}$ ▷ *Exploration*
7: If $t \in \{(q-1)L+1, \cdots, qL\} \setminus \{t_{q,1}, \cdots, t_{q,K}\}$, play $y_t = x_q$ for F_t ▷ *Exploitation*
8: **end for**
9: For $k = 1, 2, \ldots, K$, $g_{q,k} \leftarrow \frac{d}{\delta} F_{t_{q,k}}(y_{t_{q,k}}) u_{q,k}$, $d_q^{(k)} \leftarrow (1-\rho_k) d_q^{(k-1)} + \rho_k g_{q,k}$, compute $\langle v_q^{(k)}, d_q^{(k)} \rangle$ as reward for $\mathcal{E}^{(k)}$, and feed back $d_q^{(k)}$ to $\mathcal{E}^{(k)}$
10: **end for**

Bandit Submodular Set Maximization

An Impossibility Result

A natural idea: apply `Bandit-Frank-Wolfe` on F_t, the multilinear extension of the discrete objective function f_t, subject to $\mathcal{K}$, where $\mathcal{K}$ is the matroid polytope of the matroid constraint $\mathcal{I}$. Recall the one-point gradient estimator, we required the rounding scheme $\mathrm{round}_\mathcal{I}: [0,1]^d \to \mathcal{I}$ to satisfy the following unbiasedness condition

$$\mathbb{E}[f(\mathrm{round}_\mathcal{I}(x))] = F(x), \quad \forall x \in [0,1]^d$$

for any submodular set function f on the ground set Ω and its multilinear extension F.

- We showed these kind of rounding schemes does NOT exist (Lemma 2).

Responsive Bandit Submodular Maximization Problem (`RBSM`)

If $X_t \notin \mathcal{I}$, we can still observe the function value $f_t(X_t)$ as feedback, while the received reward at round t is 0 (since the subset that we play violates the constraint $\mathcal{I}$).

The description of our proposed algorithm `Responsive-Frank-Wolfe` is outlined in Alg. 3 of the paper.

References

[1] An Bian, Baharan Mirzasoleiman, Joachim M. Buhmann, and Andreas Krause. Guaranteed non-convex optimization: Submodular maximization over continuous domains. In *AISTATS*, February 2017.

[2] Lin Chen, Christopher Harshaw, Hamed Hassani, and Amin Karbasi. Projection-free online optimization with stochastic gradient: From convexity to submodularity. In *ICML*, page to appear, 2018.

[3] Lin Chen, Hamed Hassani, and Amin Karbasi. Online continuous submodular maximization. In *AISTATS*, pages 1896–1905, 2018.