We thank the reviewers for the detailed comments. We will address all the minor issues and do not discuss them individually here. We will also add high-level pictures and proof sketches in the revision.

Novelty and Originality: We would like to first describe briefly the key differences between our work and the existing methods.

(1) [17, 18] are the only prior work on `OCSM` where up to $K = T^{3/2}$ gradients are required per iteration. Note that K is not a constant, but a function of the total number of rounds T. Thus reducing K to 1 is an important step. To do so, we proposed *a series of novel methods* including *the blocking procedure* and *the permutation methods* (L114-131). Besides these, as noted in L137-144, *a novel error analysis* was performed although we relied on the same averaging technique proposed as in [37, 38].
(2) Th extension from online setting to bandit is far from trivial given the one-point estimator. Indeed, the bandit information setting is far more challenging than the online (full information) setting and several novel steps are required to design a low-regret algorithm: First, in the bandit algorithm, the point at which we play and the point at which we get the gradient estimation are different (L206-209). To circumvent this issue, we proposed *the biphasic (exploration/exploitation) method* (L210-218). Second, the point for estimation may fall out of the constraint set (L178-179). In [23], this issue was resolved by assuming that $rB^d \subset \mathcal{K} \subset RB^d$, which does not hold for many DR-continuous submodular function, whose domain is defined to be a subset of the non-negative orthant. Therefore we introduced the definition of δ-*interior*, and explained how it can help us address the bandit problem (L180-188). We also proposed a method to *construct proper* δ-*interior* (L195-197) and prove the result by a geometric analysis (Lemma 1). We established the regret bound based on Lemma 1 (Theorem 2), but also provided the *result for general constraint* (Theorem 4, Lines 219-221).
(3) By providing the hardness result (Lemma 2), we showed that it is difficult to extend our method directly from continuous settings to discrete case. Then we considered the `RBSM` model, and established a sublinear regret bound.

Response to Reviewer #1:
Q1: Using other gradient estimation techniques may lead to better regret bounds. **A1:** We agree with the reviewer and are grateful for suggesting an interesting future research direction.

Q2: Provide results for the case that the horizon T is not available offline. **A2:** Using the doubling trick (Auer et al., 1995) will easily extend our methods to the cases where T is unknown. The exploration and exploitation phases are similar to those in Alg. 2, 3.

Q3: Provide numerical experiments and compare its running time with previous algorithms for this problem. **A3:** We thank the reviewer for the suggestion. However, since our algorithms are the first with a sublinear regret bound for both one-shot online learning and continuous bandit problems, there is no previous work to compare with.

Q4: Extension to continuous submodular functions. **A4:** Our results only apply to DR-submodular functions. We tried to make it clear in the abstract and throughout the paper. We are sorry for the confusion.

Response to Reviewer #2:
Q1: Some detailed comments. **A1:** The definition of radius uses the assumption that the constraint contains 0. Assumption 5 is an analogue of the assumption in [23], which assumes that $rB^d \subset \mathcal{K} \subset RB^d$. This assumption is listed in the statement of Theorem 2 (4 *to* 6 on the first line). We assume that r is given.

Q2: Applications of `RBSM`. **A2:** In theory, `RBSM` can be regarded as a relaxation of `BSM`, which helps us to better understand the nature of `BSM`. In practice, the responsive model (not only for submodular maximization or bandit) has potentially many applications when a decision cannot be committed, while we can still get the potential outcome of the decision as feedback. For example, suppose that we have an replenishable inventory of n items. Each time there is a customer coming with a utility function unknown to us. We allocate a collection of items to her, and the goal is to maximize the total utility (reward) of all the customers. We may use a partition matroid to model diversity (in terms of category, time, *etc*). In the `RBSM` model, we cannot allocate the collection of items which violates the constraint to the customer, but we can use it as a questionnaire, and the customer will tell us the potential utility if she received those items. The feedback will help us to make better decisions in the future. Similar examples include portfolio selection: when the investment choice is too risky, *i.e.*, violates the recommended constraint set, we may stop trading and thus get no reward on that trading period, but at the same time observe the potential reward if we invested in that way. We will add more examples in the revision.

Response to Reviewer #3:
Q1: The resulting regret bounds are worse than the previous one, and shows a limitation of the techniques. **A1:** The previous $O(\sqrt{T})$ bound is achieved by using $\sqrt{T}$ exact gradients or $T^{3/2}$ stochastic gradients per round, while our method only needs one *single* gradient. Our result opens the possibility of achieving sub-linear regret with only one gradient evaluation. We agree that it is an interesting future work to achieve the same $O(\sqrt{T})$ regret bound.