We would like to thank all the reviewers for their thoughtful remarks. We will revise the manuscript to fix the typos pointed out by Reviewer 4 and implement the suggestions of Reviewer 2 regarding ways to improve the readability of the paper. In the following, we address the main concerns raised by the reviewers.

Q1: Applications and examples of continuous submodular functions beyond DR-submodularity.

A1: Maybe the simplest example of a continuous submodular function that is not DR-submodular is the quadratic function $F(x) = x^T Hx + h^T x + b$ where only the off-diagonal entries of H are non-positive (and there is no restrictions on the diagonal entries). Moreover, as we mentioned in the introduction, continuous submodular functions naturally arise as the negative log-densities of probability distributions. For instance, a distribution p on $\mathcal{X}$ is called multivariate totally positive of order 2 (MTP2) if $p(x)p(y) \leq p(x \vee y)p(x \wedge y)$ for all $x, y \in \mathcal{X} \subset \mathbb{R}$. MTP2 implies positive association between random variables. As an example, a multivariat Gaussian distibution is MTP2 if an only if its inverse covariance matrix has non-positive off-diagonal entries. Therefore, finding the the most likely configuration in this setting amounts to maximizing a continuous submodular function. Finally, as we mentioned in the paper, finding the mode of multivariate logistic, Gamma and F distributions, as well as characteristic roots of random Wishart matrices amounts to maximizing a continuous submodular function.

Q2: Why the proposed method cannot handle more complicated constrains?

A2: As mentioned in the paper, our algorithms resemble algorithms for maximizing a submodular *set* function subject to a knapsack constraint. Moreover, this is the case even when all the coordinates have a coefficient of 1 in the constraint, which corresponds to a simpler cardinality constraint. An intuitive reason for that is that we view a monotone non-DR submodular continuous function as a monotone submodular *set* function over an infinite ground set consisting of pairs (i, v), where i is a coordinate and v is the value assigned to this coordinate. Unfortunately, this intuitive reduction converts cardinality constraints into knapsack constraints, forcing us to employ knapsack techniques even when handling cardinality-like constraints. Similarly, other constraints will also become more involved if the above intuitive reduction is applied to them, usually leading to constraints that are difficult to handle.

Q3: Relationsip between the current work and the paper by Soma and Yoshida.

A3: We would like to thank Reviewers 3 and 4 for pointing out to us the paper of Soma and Yoshida, which is indeed related to ours. However, despite this relationship, it is not easy to derive a result for our setting based on the paper Soma and Yoshida for a few reasons. Perhaps the most significant of these reasons is that we believe (and they confirmed) that the result of Soma and Yoshida is not entirely correct, as is hinted by the lack of an enumeration step in their algorithm. Specifically, the main issue is that the proof of their Lemma 5 invokes Lemma 4 with $k' + \Delta(a)$, but without verifying that $k' + \Delta(a)$ is upper bounded by $k_{\max}$ (which is a necessary condition of Lemma 4). We have indeed contacted Soma and Yoshida regarding this error. Here is a part of their response: "Yuichi and I (Tasuku) discussed the issue and confirmed that our algorithm has a flaw. As you pointed out, Lemma 4 cannot be applied especially when $k_{\max} = r - y(E)$. We did not find out an easy fix".

Q4: Can the work by Soma and Yoshida be easily extended to the continuous setting?

A4: Even if the result of Soma and Yoshida was corrected, extending it to our setting is non-trivial. There are two natural ways in which one might try to achieve this goal. One way is to try to create a “black box” reduction from our continuous setting to their discrete setting by considering a fine enough lattice. This could only work if the first derivative of the objective function was bounded. However, when the first derivative can be very large (which is allowed in our setting), there might be a significant difference between the maximum objective value achievable at the *feasible* points of the lattice and the real maximum objective value. The other way in which one might try to extend the result of Soma and Yoshida to our setting is by employing ideas from this result in the creation of a new algorithm for our continuous setting. As explained in the reviews, some of our ideas overlap with those of Soma and Yoshida. However, our work deviates significantly from that of Soma and Yoshida, and cannot be viewed as an easy derivative of their work. Here are two pieces of evidence for that.

1) The time complexity of Soma and Yoshida depends on τ—the ratio between the maximum value of a solution whose support consists of a single coordinate and the minimum increase in the value of a solution when its ℓ_1 norm is increased by 1. This definition of τ is strongly connected to the discrete nature of the problem considered by Soma and Yoshida as it represents the ratio between the maximum and minimum non-zero contributions that a single coordinate can have. Extending τ's definition to our continuous domain is problematic since continuity means that a coordinate can contribute an arbitrarily low amount, leading to an infinite value for τ. Instead, we force our algorithm to increase the ℓ_1 norm of its solution by at least some minimal amount in each iteration, which requires us to develop a bound on the loss due to this extra restriction.

2) The algorithm of Soma and Yoshida heavily depends on binary searches, which are natural in their discrete setting, but make little sense in our continuous setting. This required us to develop the more sophisticated techniques represented by Propositions 3.1 and 5.1.