

1 We thank all the reviewers for their work in this challenging time. We will fix all typos and apply the clarifications
2 suggested in the reviews. Below, we address specific questions and concerns.

3 **Reviewer 1. Q1:** *In some applications of DPP, the actual features vectors are not available,[...]*

4 **A1:** Thanks for suggesting the problem. While the Online-DPP algorithm only requires computing ridge leverage
5 scores, it is not clear if the entire analysis goes through. This is an interesting point that we will verify.

6 **Q2:** *Out of curiosity I was thinking if your analysis in the first part is tight and I came up with this example.*

7 **A2:** Beautiful! Thanks for this observation :)

8 **Reviewer 2. Q1:** *Experimental focus on Online-LS and not Online-DPP*

9 **A1:** Thanks for the suggestion, we decided to focus our experiments on Online-LS because it is a more practical
10 approach to the problem. In line 148, vol_{\min} should be vol_{first} . Finally, Shuttle dataset refers to Statlog, we will clarify
11 this in the final version.

12 **Reviewer 3. Q1:** *Experiments do not seem to be convincing [...]*

13 **A1:** Thanks for raising those points. For (1), please refer to the answer to Reviewer 2. For (2), we compared with a
14 natural heuristic for the problem, it is not clear to us how to define an alternative heuristic for the online problem. An
15 additional advantage of Online-Greedy is that it is a simplified version of Online-LS and so it shows how our theoretical
16 results compare with experiments. For (3), we think that it is interesting that Online-LS (a) always seems to do strictly
17 better, (b) the gap is not too large on chosen datasets even though we show that it can be arbitrarily bad (Appendix E). It
18 suggests that the choice of algorithm in practice depends on how much accuracy we care about.

19 We will also add references to the recent works on submodular optimization.

20 **Reviewer 4. Q1:** *The proposed methodology works for small values of k only.*

21 **A1:** Our algorithm works for all k . The approximation factor depends exponentially on k but this is unavoidable for
22 this problem *even in the offline setting* unless $P=NP$ (see answer to Q6)

23 **Q2:** *It is not clear whether the output of Algorithm 2 (Online-DPP) has size k , while the is the very purpose of the paper*

24 **A2:** Algorithm 2 (Online-DPP) provides a *core-set* (a set that is guaranteed to contain a k -subset that provides a good
25 approximation). As is standard, a core-set has size larger than k . In lines 199-209, we explain how to maintain a
26 solution of size exactly k over the course of the algorithm.

27 **Q3:** *There is no comparison with streaming algorithms, nor with a random baseline[...]*

28 **A3:** Streaming algorithms do not optimize the number of times the maintained solution “changes” (which is the key
29 parameter in the online version). Our Online-Greedy and Online-LS may be viewed as variants of streaming algorithms
30 where we do not change the solution unless the value improves considerably. Moreover, we show that our solution
31 quality is comparable with the *offline* algorithm, which is much stronger than the streaming and random baselines.

32 **Q4:** *Supplementary code, and proposed change to the title.*

33 **A4:** We will make the code publicly available once the paper is accepted, and will consider the title change, thanks!

34 **Q5:** *Magnitude of the solution in the experiments approaches machine precision?*

35 **A5:** In our experiment (note that we track the change of the cost of the solution after each column insertion) we did not
36 observe any sort of numerical instability.

37 **Q6:** *How useful is the exponential $k^{O(k)}$ approximation factor [...]*

38 **A6:** An exponential dependence on k in the approximation factor is the best possible guarantee unless $P = NP$ (as
39 discussed in lines 39-40), even in the *offline* setting. This is effectively because volume in k dimensions scales as the
40 k th power. This is why some works use $vol()^{1/k}$ as the measure of interest.

41 **Q7:** *Please give the time complexity of $Tridge(kL) + Tvol(k)$.*

42 **A7:** The time complexity of $Tridge(kL)$ is discussed in lines 185-187. To compute the volume of k columns, $Tvol(k)$,
43 you need to consider the columns in an arbitrary order and find the distance of each column to the span of previous
44 ones. This can be done by maintaining an orthonormal basis of prefixes of this sequence of columns. So for the i -th
45 column ($2 \leq i \leq k$), the inner product of it with the previous $i - 1$ columns is needed. So the volume computation is
46 reduced to $O(k^2)$ inner product computations. We will clarify this in the final version of the paper.

47 We will incorporate the other proposed suggestions (including fixing the reference to Derezhinski et al.) and typo fixes to
48 improve the readability. The *DPP terminology* (instead of optimal design) is preferred mainly to be consistent with
49 recent work on coresets (which is why we mention optimal designs in related work).