# View Reviews

**Paper ID**
1073

**Paper Title**
Bayesian experimental design using regularized determinantal point processes

**Reviewer #1**

## Questions

**1. [Summary] Please summarize the main claims/contributions of the paper in your own words.**
This paper proposes an approximate algorithm for various types of Bayesian experimental design problems, i.e., finding a subset minimizing the (A/C/D/V)-optimality function. The authors observe that approximate solutions can be found by sampling from the determinantal point process (DPP) where a variety of works were proposed DPP sampling methods (e.g., Kulesza and Taskar, 2012). The main contribution is to analyze that samples from the (regularized) DPP can be bounded by the optimality functions with a proper weight (Lemma 6). Furthermore, utilizing a SDP solver (for a good weight) can lead to a tight approximate solution of the experimental design. Compared to other existing methods, the proposed algorithm allows both a smaller cardinality and reduced runtime. Empirical results show that the proposed method can find a similar or better solution than other competitors on the A-optimal design under real-world datasets.

**2. [Detailed Comments] Please enter a detailed review describing the strengths and weaknesses of the submission.**
It is interesting that the regularized DPP can be a correlation DPP with a closed-form (Lemma 2 & 3) and this can allow an efficient DPP based sampling algorithm for experimental design that guarantees a tight approximation ratio. However, this sampling framework is very similar to the proportional volume sampling (Nikolov et al., 2019) where the previous work was also based on SDP and a sampling method from determinantal distribution. The main difference of this work is to generalize a classical experimental design to the Bayesian setting with a regularization regime and propose that the volume sampling can be DPP sampling under this setting. Although this makes a drastic runtime improvement (after SDP), the proposed method still depends on SDP solver (which requires expensive computations in general) and the analysis and contribution are similar to the previous work. Furthermore, experiments do not show outperforming results other existing methods, e.g., greedy bottom-up seems to be the best method in terms of both the optimality value and runtime.

In short, the connection between the regularized DPP and the experimental design is novel, but the framework and analysis are incremental compared to the prior work. Also, empirical results are not promising so I vote this for weak reject.

**3. [Score] Please provide an overall score for the submission.**

Weak Reject: Borderline, tending to reject

**4. [Questions for the Authors] Please provide questions for authors to address during the author feedback period and point out which improvement would improve your score.**

1. Is it possible to replace the volume sampling of (Nikolov et al., 2019) with the DPP sampling algorithm (allowing some assumptions)? If possible, will the complexity be reduced?

2. Is the method of (Nikolov et al., 2019) possible to guarantee the same results for C- and V-optimality? It seems to be possible since they are linear functions of (\Sigma+A)^{-1}.

3. What is the runtime complexity of greedy bottom-up? Is it possible to compare their runtimes quantitatively? And it would be better to provide the overall complexity of the proposed algorithm including SDP solver (with the number of iterations).

4. Experiments only report trivial algorithms (except greedy bottom-up). Do authors have any chance to compare the results to algorithms with an SDP solver? (i.e., methods in Table 1)

5. In experiments, the regularizer matrix A is chosen by a scaled identity matrix. Are there any other options for A in practice?

6. In proof of Corollary 1, the fourth equality should be modified to “<= det( \sum_i p_i x_i x_i^\top)^{-1/d}”.

**5. [Review Summary] Please enter a 1-2 sentence summary of your review explaining your overall score.**

The key observation that Bayesian experimental design is related to the determinantal point process is novel, but analysis and the proposed algorithm are incremental. Experimental results are also not promising compared to other existing methods.

**6. [Reproducibility] Are the experiments (if present) detailed enough to allow reproducibility?**

All experimental settings are clear and are possible to reproduce.

**7. [Code] Was the code made available by the authors?**

No

**8. [Expertise] Please rate your expertise on the topic of this submission.**

High: Reviewer has published on the topic.

**9. [Confidence] Please rate your confidence in the score assigned.**

High: Reviewer has understood the main arguments in the paper, and has made high level checks of the proofs.

**Reviewer #2**

## Questions

**1. [Summary] Please summarize the main claims/contributions of the paper in your own words.**

This paper provides polynomial time algorithms for Bayesian experimental design which works for smaller values of k; namely, when k depends on an effective dimension of the data covariance. At the heart of these new algorithms is a sampling procedure applied to the SDP relaxation. The sampling procedure relies on a newly introduced regularized determinantal point process, basic aspects of which are investigated in the paper and may be of independent interest. Finally, empirical results confirm that the rounding is competitive against existing methods.

**2. [Detailed Comments] Please enter a detailed review describing the strengths and weaknesses of the submission.**

This was a wonderful paper to read because it is organized very well and the technical exposition is very clear. The clear exposition is one major strength of the paper. Authors are clear to outline prior work on experimental design, especially with regards to (1) what two types of questions are typically asked and (2) prior dependence on d, rather than effective dimension d. This insight of effective dimension is quite nice and seems to get very tight results, from what experiments show.

Another strength of the paper is the introduction and analysis of the regularized determinantal point processes, which is likely of independent interest outside of experimental design. In particular, Theorem 3 provides a nice equivalent description of the regularized determinantal point process as a low-rank DPP with Bernoulli random variables. Moreover, lemmas 4 and 5 work out some simple properties of the regularized DPP. The initial investigation into these regularized DPPs in this paper will likely find uses elsewhere.

I have a few minor comments towards the paper, but they are all positive and I wouldn't describe them necessarily as weaknesses.

Authors reported that the greedy algorithm features the same run time as their proposed rounding method + SDP. For the greedy algorithm, are marginal gains computed by explicitly computing a matrix inverse / linear system solve at each step? If so, then the greedy algorithm may be greatly sped up, as discussed in [Harshaw et al, ICML 2019], what is Section 5.1 in the arXiv version. The idea is to use matrix inversion lemma to quickly evaluate functions and maintain inversions. This speeds up the greedy algorithm by several orders of magnitude in practice. It's fine if authors weren't able to compare this much faster runtime, but I would recommend making a note that faster greedy implementations are possible. Of course, they are less desirable in the sense that they do not provide as strong an approximation guarantee in the worst case.

Minor Comments:

1. In the introduction, plain $\Sigma$ is used for the covariance of the distribution of linear function w given outcomes y_S, but later the subscripted $\Sigma_{w \mid y_S}$ is used. I recommend defining the posterior covariance using the same notation so as to avoid confusion, especially with the $\Sigma$ appearing in

function definitions (1) - (4).
2. A question on the posterior covariance: I guess that inverse is only well defined when X_S^T X_S + A is invertible. If A is not invertible, then we might require X_S to have full column rank, right? This seems like this won't happen when A = 0 and |S| = 1, but perhaps I'm mistaken. So, it seems that for A=0, one won't be able to run the greedy algorithm initialized at the empty set since the function isn't defined here. Can you speak more to this minor point?
3. Question 2 reads "what is the upper bound on OPT_k"? but you might wish to use the word "an upper bound" rather than "the".
4. Definition 1 ends with the sentence "We will use the shorthand d_A when referring to $d_A(\Sigma_X)$", but this seems more of a notational point and not part of the definition, so I would recommend removing it from the definition block and moving it to the main text.

**3. [Score] Please provide an overall score for the submission.**
Strong Accept: Outstanding paper

**4. [Questions for the Authors] Please provide questions for authors to address during the author feedback period and point out which improvement would improve your score.**
Did authors run the naive implementation of the greedy algorithm or a faster implementation, e.g. as described in [Harshaw et al, ICML 2019]? As I said, the naive implementation is fine, but it'd make for a fairer comparison to state that faster implementations are possible.

**5. [Review Summary] Please enter a 1-2 sentence summary of your review explaining your overall score.**
I think this is a very strong submission, both with respect to the improved Bayesian Experimental design results but also with respect to the newly introduced regularized DPPs.

**6. [Reproducibility] Are the experiments (if present) detailed enough to allow reproducibility?**
Yes. However, since running times are informally discussed, it might be a good idea to include the computing infrastructure used.

**7. [Code] Was the code made available by the authors?**
No

**8. [Expertise] Please rate your expertise on the topic of this submission.**
High: Reviewer has published on the topic.

**9. [Confidence] Please rate your confidence in the score assigned.**
High: Reviewer has understood the main arguments in the paper, and has made high level checks of the proofs.

**Reviewer #3**

## Questions

**1. [Summary] Please summarize the main claims/contributions of the paper in your own words.**
The authors consider the Bayesian experimental design problem: given a set of experiments x_1, ..., x_n,

choose a set of size k such that minimize some optimality criteria. There are many different optimality criteria. In this work the authors consider A, C, D, and V optimality.

The problem now becomes a combinatorial optimization problem to choose the best set of experiments to perform. The authors develop efficient approximation guarantees when the number of experiments k ~= d_A, where d_A is the reduced dimension and is lower than the full dimension d due to the Bayesian prior. Prior work was not in the Bayesian setting and was only able to guarantee approximations when k ~= d. Their approach is based on sampling from a new class of distributions they call regularized DPP.

Additionally, the authors consider

**2. [Detailed Comments] Please enter a detailed review describing the strengths and weaknesses of the submission.**

I found this to be an interesting paper. The new class of DPPs is interesting and they provide an efficient way to sample from it. Additionally, their results in the Bayesian setting allow guarantees for more values of k.

I have a question on the sampling distribution for Theorem 1 compared to Nikolov et al, 2019. Looking at Lemma 3.2 of (https://arxiv.org/pdf/1802.08318.pdf), it seems very similar to a regularized DPP when A = 0, while the text says that it is unregularized. Besides the new approach to sample from regularized DPPs, are the algorithms essentially the same when A = 0? Is the difference only the new way to sample?

The experiments are the weakest part of the paper. The authors only consider a single prior and optimality condition. Additionally it would be nice to see empirical calculations on the effective dimension for real datasets, as well as see the performance of the solver when k is about the effective dimension.

**3. [Score] Please provide an overall score for the submission.**

Accept: Good paper

**4. [Questions for the Authors] Please provide questions for authors to address during the author feedback period and point out which improvement would improve your score.**

Can you comment more on the runtime cost including the SDP? Specifically, in Table 1 the runtime cost is after SDP. Do all this algorithms run the same (or similar) SDP?

Can you comment more on how similar your approach is to Nikolov et al., 2019 when A = 0 and what causes the difference in runtime?

**5. [Review Summary] Please enter a 1-2 sentence summary of your review explaining your overall score.**

This paper has several interesting theoretic results. Experiments are weak but sufficient.

**6. [Reproducibility] Are the experiments (if present) detailed enough to allow reproducibility?**

Yes.

**7. [Code] Was the code made available by the authors?**

No

**8. [Expertise] Please rate your expertise on the topic of this submission.**

High: Reviewer has published on the topic.

**9. [Confidence] Please rate your confidence in the score assigned.**

High: Reviewer has understood the main arguments in the paper, and has made high level checks of the proofs.