Referee Report on "Solving Maximum-Entropy Sampling Problems Using Factored Masks"

Authors: Samuel Burer & Jon Lee

June 24, 2005.

The paper presents a branch-and-bound based algorithm for the maximum-entropy sampling problem that solves the masked spectral bound at each node using a gradient descent algorithm. The authors provide computational evidence of the benefit of their approach by comparing it to fixed mask approaches of eigenvalue bound, diagonal bound and one-big partition bound.
It is clear that enormous amount of computational work was put into this paper and the experimentations are very thorough. However, the significance of its contribution compared to its cited papers and its relevance to the *Mathematical Programming* journal may be questionable. There are some points in the paper where the authors hypothesize the behavior of their gradient descent methods without any theoretical nor empirical justification. This is not meant to put down the quality of the work. This referee would be in favor of the paper's acceptance if the editor finds it appropriate for this particular journal.

Questions & Comments

The following are some questions and comments. Some of the points may have been addressed by the authors but were not sufficiently clear for this referee.

- I was surprised that the two steepest descent approaches worked as well as it did on this function. There are some intuitive reasonings for this in the third paragraph of page 2, but it would be nice to get more insight (theoretical and/or empirical) to the algorithms' behaviors to the spectral bound function.

- How is V in the descent algorithms randomly initiated? Is it generated from some distribution? I would think there are some *a priori* knowledge about the V that can be exploited instead of using blind random generation.

- In Section 3.1, the authors state that both Algorithm 1 and 2 behave similarly across all instances of (n, s, k) on both the original and complimentary problems. Figure 3 shows just one of those instances. I would have thought that there may be convergence and running time differences for different s. Is that not the case?

- In Section 3.1, the authors provide a hypothesis to Algorithm 2's superior performance. Again, can the authors give some (it does not need to be extensive) evidence (theoretical and/or empirical) to back this up?

- Is the entropy gap a relative gap or an absolute gap? The difference in the gap behavior between the original and complementary problem is confusing in Figures 3, 6 and 7. I may be misunderstanding these illustrations but the convergence of the complementary problem looks consistently more stable.

- Figure 4 seems to show that the relative performance between Algorithm 1 and AS is somewhat dependent on s. Is that just a coincidence? Also, Figure 5 shows Algorithm 1 strictly

outperforming Algorithm AS – is this behavior different for 5 steps? If the methods are going to be used for at most 2–5 iterations in the branch-and-bound scheme, it may be more relevant to compare its short term behavior.

- In most of the paper, the actual CPU times of the algorithms are not reported since the running times are similar across Algorithm 1, 2 and AS. However, this referee would be interested in knowing the total running time to get a general feel of the difficulty of these problems.
- In Section 3.3, the authors indicate that the lower rank was not ultimately used since the total running was dominated by the eigensystem calculations. Were some heuristics for choosing k tested (i.e., have k be some function of n and s)? Also, how was the eigensystem solved in this paper?
- In Section 4, the authors very briefly state that a greedy heuristic was used to find a global lower-bound. How were the elements in N chosen greedily? I would think that a good global lower-bound would significantly reduce the size of the branch-and-bound tree. Have other heuristics been tried as well?
- This referee has no intuition why subproblems with the same (f, u) would behave similarly in terms of the original or complementary bound. Could the authors provide some intuition?
- In the "Optimization bounding strategy", the authors used quadratic interpolation to predict the bound value for the next three iterations. Is this a common techniques? And how did the authors find its effectiveness in practice?
- What is meant by "true warm start" in this context? In the simulated warm starting, is the values of V updated after branching? For example, if $i \in N \setminus U \setminus F$ is not selected for S, is the "corresponding" row and column deleted from V?
- It is peculiar that Algorithm 1 performed better than Algorithm 2, given the latter's superior performances reported in previous sections. The authors do not give much justification for this – is it due to the numerical instability arising from warm starting? Have they compared Algorithm 1 and 2 using random starting points for each node? I ask this to double check for errors in the code or implementation of the two algorithms.

Minor Comments

The following are very minor comments, mainly pertaining to language.

- (page 1) I do not believe "Branch-&-Bound" should be capitalized.
- (page 1) There should be a comma after "Section 1" in the last paragraph.
- (page 2) "can not" should be "cannot" in the second to last paragraph.
- (page 4) For consistency, it should be "$k_{max} := 2$" and "$\beta := 1/3$" in the paragraph starting "Anstreicher and Lee".
- (page 5) There should be an "is" after " One hypothesis as to why we see this behavior with Algorithm 2" in the second to last paragraph.

- (page 6) In the captions of Figure 3, “complementary” is misspelled and the “and” before it should not be italicized.
- (page 9) In captions of Figure 6, “complementary” is misspelled.