

Response to Referee Comments on the Paper:

A general approach for incremental approximation and hierarchical clustering

Guolong Lin

Chandrashekhara Nagarajan

Rajmohan Rajaraman

David Williamson

February 25, 2009

1 Overview

This document describes the revisions made to the paper in response to the comments of the referees. Before proceeding any further, we would like to thank the referees for the time and energy they have spent studying our paper thus far.

We have made a number of revisions to the paper to address the referee comments. We have made significant changes to the presentation in Sections 4 and 5. Furthermore, we have made the corrections and have added explanations, as suggested by the referees. The revisions are discussed in detail in the following two sections (we have arbitrarily numbered the referees 1 and 2).

2 Response to the comments of Referee 1

2.1 Response to high-level comments

1. **Comment:** Choice of problems.

Response: We would like to keep the k -vertex cover and k -set cover problems since there is quite some prior work on the problems, and they illustrate the application of our framework to a wider context. We believe these problems are also well-motivated and have mentioned the applications in the related work section, along with citations to prior work on these problems.

2. **Comment:** Definition of model.

Response: We agree with you that the current definition may make the median problems more difficult to follow since the ordering is the opposite of what is arguably more natural. We feel, however, that reversing the ordering would have a similar effect on the other problems, since the parameter k (or equivalent) in those problems is a lower bound as opposed to an upper bound. In the end, we had to make a choice (since it is more preferable than having two sets of definitions) and have used the current one. We have added a footnote in the framework section that explains our reasoning.

3. **Comment:** k -medians versus facility location.

Response: We have added a remark in the facility location section, explaining why the two orderings are different.

4. **Comment:** Three variants of the algorithm.

Response: The second variant of the algorithm `ALTINCAPPROX` gives improved approximations for the hierarchical median and means problems. As currently defined, however, `ALTINCAPPROX` does not apply to the incremental facility location problem. While it may be possible to rework the definition, we feel that the first variant `INCAPPROX` seems quite natural and worthwhile to include for this purpose. We believe that `ALTINCAPPROX` provides a useful alternative way of presenting the incremental framework. While the resulting improved competitive ratios are not small (as pointed out by the referee), there is the potential for further improvements in the results if better offline approximation ratios are obtained. Finally, the third variant `BOUNDEDINCAPPROX` is based on new concepts, so seems worthy of inclusion.

2.2 Response to specific comments

The numbering of the items below is the same as that of the comments in the referee report.

1. Done.
2. We have added a paragraph in the related work section that provides motivation for the k -vertex cover and k -set cover problems. We have also added citations to other papers that have considered the two problems.
3. We have replaced “ n clusters” with “a cluster for each point”.
4. We have added one citation for each of the problems.
5. We have illustrated the notation through the k -MST and k -median problems.
6. We have changed the assumption to the following: the cost of any feasible solution is either zero or bounded away from zero; i.e., $\inf\{\text{cost}(S) : S \in U, \text{cost}(S) > 0\} > 0$. The cost scaling has been moved into the proof.

The above assumption is needed in the proof, and is trivially satisfied whenever the set of feasible solutions is finite. It may not always hold, however, when the set of feasible solutions is infinite; for example, this is the case in the incremental facility location problem and we address it in the relevant section. We have pointed this out in the remark immediately after the proof of Theorem 2.1.

7. We have fixed the subscripts in the figure.
8. Done.
9. Corrected.
10. We have written all the inline fractions in the a/b form, except for a small number that had several terms each and seemed to deserve the $\frac{a}{b}$ form.
11. Corrected.
12. We have added two figures and more explanation motivating bounded envelopes.

3 Response to the comments of Referee 2

- **Comment:** In a number of places where fractions are used, there is some sort of typesetting problem that makes the division bar of the fraction invisible.

Response: Hopefully, you do not encounter this problem in the revision.

- **Comment:** In Section 2.1, I think the poset terminology (e.g. “maximal chain”) should be defined for clarity.

Response: We have added the definitions of chain and maximal chain at the places where they are first used.

- **Comment:** In Section 3.3, the notation for the set of reals is not consistent.

Response: Corrected.

- **Comment:** In Lemma 5.1, I believe that the set D should in fact be the set C throughout the proof.

Response: Corrected.

- **Comment:** Section 4 could benefit with an introduction that gives motivation for and examples of “bounded envelope” problems.

Response: We have added two figures and more explanation for motivating bounded envelopes.