

# INFO-F420 - Computational Geometry - Project

## Geometric Approximations via Coresets

Charles HAMESSE (École Polytechnique)

November 2016

### Abstract

The paradigm of coresets has recently emerged as a powerful tool for efficiently approximating various extent measures of a point set  $P$ . Using this paradigm, one quickly computes a small subset  $Q$  of  $P$ , called a coreset, that approximates the original set  $P$  and then solves the problem on  $Q$  using a relatively inefficient algorithm. The solution for  $Q$  is then translated to an approximate solution to the original point set  $P$ .

## 1 Proposal

I would like to work on coresets and implement concrete coreset construction algorithms. In fact, the literature on this topic remains quite theoretical and lacks of actual examples to me.

Coresets (also called  $\varepsilon$ -approximations) are of great interest regarding today's use of technologies: we deal with huge amounts of data, coming from high-throughput sources. For example, in computer vision or machine learning, we deal with  $n$ -dimensional data points and often implement border (or *extent*) recognition techniques to reveal the shape, the boundaries of some objects. Hence being able to approximate the extent of such point sets seems a valuable skill. In this sense, I will first consider problems of shape fitting using coresets. That is, given an input point set  $P$ , the goal is to find the best fitting shape from some class to  $P$ . Just to name a few, these shape classes include for example squares, circles as well as cubes, spheres or even cylinders.

**Interactive experience** I will put emphasis on the visual aspect. As stated above, it seems like there isn't much visual educational material on this type of geometric approximation. This leads to the idea for the main interactive part of the project: implementing a shape fitting algorithm based on various 2D or 3D point sets, given the base shape classes. Points sets can come from various sources, e.g. user input (if 2D), random generation or generation from a base shape after some alteration (e.g. generating a cylinder then adding noise to its points' coordinates).

**Theoretical background** I would implement the technique described in *Approximating Extent Measures of Points* [2] for the task. I quickly went through this paper and a lot of related work, and it seems a good starting point. More precisely, I already have an idea of what topics I *must* talk about in the theoretical part of the project:

1. Introduction:
  - (a) Definition of coresets and critical notions
  - (b) Description of shape fitting problems (e.g. sphere-shell problems or cylindrical-shell problems)
2. Preliminaries
  - (a) Envelopes and extent
  - (b) Directions and directional width

3. Approximating the extent of linear functions
  - (a) Reduction to fat point set (as in Lemma 3.1, briefly)
  - (b) Weaker bound on  $\varepsilon$ -approximation (as in Lemma 3.4, briefly)
  - (c)  $\varepsilon$ -approximation construction (as in Lemma 3.5, important)
  - (d) Other subjects (as in Lemma 3.6 to 3.11, maybe briefly)
4. Applications
  - (a) Approximating faithful extent measures
  - (b) Minimum-width spherical shell
  - (c) Minimum-width cylindrical shell

I think this is a proper way to introduce the concept of coresets and play with it, but it is of course subject to potential changes along the way.

## References

- [1] Pankaj K. Agarwal, Sariel Har-Peled, Kasturi R. Varadarajan, Geometric Approximation via Coresets: <http://sarielhp.org/papers/04/survey/survey.pdf>
- [2] Pankaj K. Agarwal, Sariel Har-Peled, Kasturi R. Varadarajan, Approximating Extent Measures of Points: <http://sarielhp.org/p/01/fitting/fitting.pdf>
- [3] Jeff M. Phillips, Coresets and Sketches: <https://arxiv.org/pdf/1601.00617v2.pdf>
- [4] Gill Barequet, Sariel Har-Peled, Efficiently Approximating the Minimum-Volume Bounding Box of a Point Set in Three Dimensions: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.78.229>
- [5] Dan Feldman, Micha Feigin and Nir Sochen: Learning Big (Image) Data via Coresets for Dictionaries: <https://people.csail.mit.edu/dannyf/JMIV.pdf>
- [6] Sariel Har-Peled, Geometric Approximation Algorithms (course from Stanford University): <https://graphics.stanford.edu/courses/cs468-06-fall/Papers/01%20har-peled%20notes.pdf>
- [7] Jian Zhang: Massive Data Streams in Graph Theory and Computational Geometry, :<http://www.cs.yale.edu/homes/jf/Jian-thesis.pdf>