On Clustering Induced Voronoi Diagrams

Danny Z. Chen1  Ziyun Huang2  Yangwei Liu2  Jinhui Xu2

1 Department of Computer Science and Engineering
University of Notre Dame
dchen@cse.nd.edu
2 Department of Computer Science and Engineering
State University of New York at Buffalo
{ziyunhua, yangweil, jinhui}@buffalo.edu

1 Overview

Voronoi diagram is a fundamental geometric structure with numerous applications in many different areas [1–3]. Ordinary Voronoi diagram is a partition of the space \( \mathbb{R}^d \) into a set of cells induced by a set \( P \) of points (or other types of objects) called sites, where each cell \( c_i \) of the diagram is the union of all points in \( \mathbb{R}^d \) which have a closer (or farther) distance to a site \( p_i \in P \) than to any other sites. In some sense, cells in a Voronoi diagram can be viewed as formed by competitions among all sites in \( \mathbb{R}^d \) such that the winner site for any point \( q \in \mathbb{R}^d \) is the one having a larger “influence” on \( q \) defined by its distance to \( q \).

In this paper, we generalize the concept of Voronoi diagram to Clustering Induced Voronoi Diagram (CIVD). In CIVD, we consider a set \( P \) of \( n \) points (or other types of objects) and a non-negative influence function \( F \) which measures the joint influence \( F(C, q) \) from each subset \( C \) of \( P \) to any point \( q \) in \( \mathbb{R}^d \). The Voronoi cell of \( C \) is the union of all points in \( \mathbb{R}^d \) which receive a larger influence from \( C \) than from any other subset \( C' \subseteq P \). This means that CIVD considers all subsets in the power set \( U = 2^P \) of \( P \) as its sites (called cluster sites), and partitions \( \mathbb{R}^d \) according to their influences. While CIVD in general can have exponentially many cells, it is possible that for some interesting influence functions only a small number of subsets in \( U \) have non-empty Voronoi cells, making the problem solvable.

Relation to Previous Works: To our best knowledge, there is no previous work on the general CIVD problem. Our CIVD model obviously extends the ordinary Voronoi diagrams [2], where each site is a one-point cluster. (Note that the ordinary Voronoi diagrams can be viewed as special CIVDs equipped with proper influence functions.) Some Voronoi diagrams [3, 12] allow a site to contain multiple points, such as the \( k \)-th order Voronoi diagram [3]. Some two-point site Voronoi diagrams were also studied [4, 5, 7, 8, 10, 11, 13], in which each site has exactly two points. Obviously, such Voronoi diagrams are different from CIVD.

For Vector CIVD, influence between any two points \( p \) and \( q \) is defined by a force-like vector. The problem is related to the N-body problem [9], which shares with the Vector CIVD problem a similar idea of modeling joint force by influence functions. Density-based CIVD enables us to generate all density-based clusters as well as their approximate Voronoi cells. The problem is related to density-based clustering which is widely used in many applications.

1.1 Results and Techniques

The main result of the paper is a general technique called Approximate Influence(AI) Decomposition, which can be used to generate \((1 - \epsilon)\)-approximate CIVD. We also apply AI decomposition to develop assignment algorithms for vector CIVD and density base CIVD. Below is a list of our main results.

* A preliminary version of this work has appeared in the Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2013)
Properties of Influence Function: We investigate the general and sufficient conditions which allow the influence function to yield only a small number of non-empty approximate Voronoi cells. We show that the following three properties are sufficient: Similarity Invariant property, Locality property, Local Domination property. The first property means that for any point \( q \in \mathbb{R}^d \), its maximum influence cluster site remains the same after a similarity transformation about it. The second property indicates that a small perturbation on a cluster site \( C \) or \( q \) only changes slightly their influence. The third property implies that each cluster site may have dominating influence in its neighborhood.

Approximate Influence(AI) Decomposition: We present a standalone technique called approximate influence decomposition (or AI decomposition) for general CIVD problems. In \( O(n \log n) \) time, this technique partitions the space \( \mathbb{R}^d \) into a nearly linear number (i.e., \( O(n \log n) \)) of cells so that for each such cell \( c \), there exists a (possibly unknown) subset \( C \subseteq P \) whose influence to any point \( q \in c \) is within a \((1 - \epsilon)\)-approximation of the maximum influence that \( q \) can receive from any subset of \( P \), where \( \epsilon > 0 \) is a fixed small constant. In this technique, we also develop a new data structure called box-clustering tree, based on an extended quad-tree decomposition and guided by a distance-tree built from the well-separated pair decomposition [6]. In some sense, our AI decomposition may be viewed as a generalization of the well-separated pair decomposition.

Assignment Algorithms for Vector CIVD and Density-base CIVD: Based on the AI Decomposition, we develop assignment algorithms for the Vector CIVD and the Density-based CIVD problems. Particularly, we show that it is possible to determine a proper cluster site for each cell in the decomposition and form a \((1 - \epsilon)\)-approximate CIVD for each problem for any given small constant \( \epsilon > 0 \). For Vector CIVD, the assignment algorithm is based on several new techniques such as aggregation-tree and majority path decomposition, and runs in \( O(n \log^d n) \) time. For Density-based CIVD, the assignment algorithm takes \( O(n \log^2 n) \) time and can be obtained from a modification of the AI decomposition.

References