Algorithm Seminar 10/19

Abstract

We discuss the problem of attaching geometric objects at given sites on a simple curve in the plane such that none of the attachments intersect. We present a polynomial time algorithm for the case when there are at most two choices for the objects to be attached at every site. We then post several generalizations of this problem, and summarize the different approaches suggested.

1 The Simple Version : Problem Statement and Algorithm

1.1 Introduction

The motivation for this problem comes from protein folding, where one is given the "backbone" of a protein (a simple curve) along with a set of "sites" on this backbone. For every site, there is a (known) set of amino-acids that can be attached to the backbone at that site. Every site must have exactly one attachment on it. A (known) energy functional allots an energy value for every possible configuration of amino-acids attached at the given sites. This energy functional has three terms: the energy of the underlying backbone, the energy from the interaction between the attached amino-acids and the backbone, and the energy from the pairwise interaction between attached amino-acids.

The goal is to find the configuration that minimizes this energy. The above problem is known to be NP-Hard, and we relax the problem first to find a configuration of attachments which

- Do not intersect the backbone.
- Do not intersect with each other.

1.2 Problem Statement

We formalize the simple version of the problem as follows. Let $\mathcal{B}$ be a simple polygonal chain, denoted as $(b_1, b_2, ..., b_\ell)$, where b_i are the vertices of this chain. We will refer to $\mathcal{B}$ as the "backbone". Let $\mathcal{S} = \{s_i\}_{i=1}^n$ be a subset of the vertices of $\mathcal{B}$, referred to as "sites". For every site s_i there is a set S_i containing different polygonal chains that can be attached to $\mathcal{B}$ at s_i (assume that the polygonal chains have a marked point, which will coincide with s_i after the attachment). The input to the problem is $(\mathcal{B}, \mathcal{S}, \{S_i\}_{i=1}^n)$.

The goal is to find one way of attaching the polygonal chains on sites such that no two polygonal chains intersect, and no polygonal chain intersects with $\mathcal{B}$. Formally, the output is a configuration π with $\pi(i) \in S_i$ such that

- $\pi(i) \cap \mathcal{B} = \varnothing$ for all i.
- $\pi(i) \cap \pi(j) = \varnothing$ for all pairs (i, j).

Conjecture 1. *The above problem (call it Version 0) is NP-Hard if $\#S_i > 2$ for all i (proof needed, reduction to 3-SAT ?).*

We now present a polynomial time algorithm for the above problem when $\#S_i \leq 2$.

1.3 Algorithm

In a linear scan, for every S_i, one can determine which elements of S_i intersect $\mathcal{B}$ and discard those elements as they can never be in an allowed configuration. At the end of this linear scan, $\#S_i \leq 2$ for all i.

For a site s_i, denote the (at most) two allowed attachments as $S_i = \{p_{i1}, p_{i2}\}$. Let X_i be a binary variable which is 0 if $\pi(i) = p_{i1}$ and 1 if $\pi(i) = p_{i2}$. We will now use 2-SAT [1] to solve the problem of finding non-intersecting attachments.

For every pair of sites (s_i, s_j), the ordered pair (X_i, X_j) takes at most four values (00,01,10 and 11). Given one such value, one can find out whether the two attachments at sites s_i and s_j corresponding to X_i and X_j intersect or not. For every disallowed configuration (X_i, X_j), form a clause, complementing the variables that are 1 in the disallowed configuration, e.g. if $X_i = 1$ and $X_j = 0$ is disallowed (meaning that p_{i2} and p_{j1} intersect), form the clause $\overline{X_i} \vee X_j$. Then take the AND of all such clauses over all pairs. Note that every pair (i, j) gives rise to at most 4 clauses, and therefore the total number of clauses is bounded by $4\binom{n}{2} = O(n^2)$.

We now have an instance of 2-SAT. It is easily seen that the instance is valid iff there exists a valid configuration $\pi(i)$. Moreover, the linear time algorithm for 2-SAT ([1], [3]) gives as output one such valid assignment, if it exists. Hence the algorithm runs in time $O(n^2)$.

Corollary 2. *The above problem is a slight generalization of what was discussed in the seminar. Let $\mathcal{B}$ be a polygonal chain on a grid $\mathcal{G}$ (so all the edges and vertices of the chain are contained in the set of edges and vertices of the grid). Let s_i be the set of sites. For every site s_i which is not the start/end vertwx of $\mathcal{B}$, there are at most 2 neighbors of s_i in $\mathcal{G}$ which do not lie on $\mathcal{B}$. If the start (end) vertex of $\mathcal{B}$ is a site, choose any 2 of the at most 3 available attachments. We will run the algorithm on all possible 9 such combinations.*

Denote the neighbors as s_{i1} and s_{i2} and let $p_{ij} := \{\text{the edge}(s_i - s_{ij})\}$. Let $S_i := \{p_{i1}, p_{i2}\}$. Thus the only two allowed attachments are just edges of length 1 at the sites.

This problem will be denoted as the 2-D grid version of the above problem (Version 2) and is a special case of Version 1 discussed above.

2 Generalizations

We now propose several generalizations of the above problem. We divide them into two cases: one in which we have to choose from at most two given attachments per site and the other in which there might be more than two attachments to choose from.

2.1 $\#S_i \leq 2$

Version 3: Assume the same setting as in Version 1. If a configuration such that there are no mutual pairwise intersections does not exist, one can ask to minimize the number of intersections.

Conjecture 3. *The above version is NP-Hard. The way we solve Version 1 shows that this problem is related to MAX 2-SAT [1], which is known to be NP-Complete, and there is currently a* $0.940..$ *approximation, which is close to the best achievable. One can (hopefully) use this approximation algorithm in a fairly straightforward way to solve this version.**Question:** Can one do better ?*

Version 4: Assume the same setting as Version 1. One is also given the set of weights $W_i = \{w_{i1}, w_{i2}\}$ which describes the preference between attachments p_{i1} and p_{i2} at site i. One is then required to find an assignment which maximizes the total weight (subject to all the attachments being non-intersecting). One can also remove the non-intersecting requirement and allow at most k intersections and then be asked for maximizing weight.

In the simple case when $w_{i1} = 1$ and $w_{i2} = 0$, the problem seems related to weighted 2-SAT [1] which is known to be hard (it is NP-Complete), even to approximate (related to the vertex cover). One can also pose this problem as an Integer Programming problem.

Conjecture 4. *The above Version is NP-Complete. Methods from Integer programming might help.*

2.2 $\#S_i > 2$

Version 5: Assume the setting as in Version 2 (the grid version), but this time the problem is in 3-D,i.e. $\mathcal{B} \subset \mathbb{R}^3$. Now there could be 4 grid neighbors of every s_i that do not lie on the backbone, and so 4 possible attachments at this site. It is not clear that this problem can be reduced to 2-SAT, and therefore whether a polynomial time algorithm exists.

Version 6: This is just the general version in $\mathbb{R}^3$. If the general version in $\mathbb{R}^2$ is NP-Hard, this is clearly hard too.

Version 7: In this case $\mathcal{B}$ is a tree instead of a polygonal chain.

3 Related Literature

Apart from 2-SAT and Integer programming which clearly are related to the above problem, *map labeling* algorithms might also be useful. In [2] and references therein, the authors consider the problem of putting rectangles/squares (representing the text) next to points (representing landmarks) on a map in such a way that the rectangles do not overlap.

References

[1] http://en.wikipedia.org/wiki/2-satisfiability.

[2] S. Doddi, M. V. Marathe, A. Mirzaian, B. M. E. Moret, and B. Zhu. Map labeling and its generalizations. In *Proc. 8th ACM-SIAM Symp. Discrete Algorithms (SODA)*, pages 143–157, 1997.

[3] M. R. Krom. The decision problem for a class of first-order formulas in which all disjunctions are binary. In *Zeitschrift fr Mathematische Logik und Grundlagen der Mathematik*, pages 15–20, 1967.