University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Katarzyna Kowalska
Student no. 371053

Approximation and Parametrized Algorithms for Geometric Set Cover

Master's thesis
in COMPUTER SCIENCE

Supervisor:
dr Michał Pilipczuk
Instytut Informatyki

June 2020

Supervisor's statement

Hereby I confirm that the presented thesis was prepared under my supervision and that it fulfils the requirements for the degree of Master of Computer Science.

Date Supervisor's signature

Author's statement

Hereby I declare that the presented thesis was prepared by me and none of its contents was obtained by means that are against the law.

The thesis has never before been a subject of any procedure of obtaining an academic degree.

Moreover, I declare that the present version of the thesis is identical to the attached electronic version.

Date Author's signature

Abstract

W pracy przedstawiono prototypową implementację blabalizatora różnicowego bazującą na teorii fetorów σ-ρ profesora Fifaka. Wykorzystanie teorii Fifaka daje wreszcie możliwość efektywnego wykonania blabalizy numerycznej. Fakt ten stanowi przełom technologiczny, którego konsekwencje trudno z góry przewidzieć.

Keywords

blabaliza różnicowa, fetory σ-ρ, fooizm, blarbarucja, blaba, fetoryka, baleronik

Thesis domain (Socrates-Erasmus subject area codes)

11.3 Informatyka

Subject classification

D. Software
D.127. Blabalgorithms
D.127.6. Numerical blabalysis

Tytuł pracy w języku polskim

Algorytmy parametryzowania i trudność aproksymacji problemu pokrywania zbiorów na płaszczyźnie

Contents

Chapter 1

Introduction

This is some very interesting NP-completneess and APX-hardness results about geometric set cover, because the problem is cool.

Our contribution. In this work, we prove that unweighted geometric set cover with segments is fixed parameter tractable (FPT).

Moreover, we show that geometric set cover with segments is APX-complete even for unweighted axis-parallel segments, and with $1/2$-extension. Therefore, in the efficient polynomial-time approximation scheme (EPTAS) for *fat polygons* by [Har-Peled and Lee, 2009], the assumption about polygons being fat is necessary.

Finally, we show that geometric set cover with weighted segments in 3 directions is W[1]-complete. However, geometric set cover with weighted segments is FPT if we allow δ-extension.

Chapter 2

Definitions

Some definitions what geometric set cover is. $\mathcal{P}$ – set of objects, $\mathcal{C}$ – set of points. Choose $\mathcal{R} \subset \mathcal{P}$ such that every point in $\mathcal{C}$ is inside some element from $\mathcal{R}$ and $|\mathcal{R}|$ is minimal.

In parametrized setting we only look among $|\mathcal{R}| \leq k$. In weighted settings there is some $f : \mathcal{P} - > \mathbb{R}$ and we minimize $\sum_{R \in \mathcal{R}} f(R)$.

Chapter 3

Geometric Set Cover with segments

3.1. FPT for segments

3.1.1. Segments parallel to one of the axis

You can find this in Platypus book.

We'll show $\mathcal{O}(2^k)$ branching algorithm. Let's take point K that hasn't been covered yet with the smallest coordinate in lexicograpical order. We need to cover K with some of the remaining segments.

We choose one of the 2 directions on which we will cover this point. In this direction we take greedly the segment that will cover the most points (there are points in $\mathcal{C}$ only on one side of K in this direction, so all segments covering K in this direction create monotone sequence of sets – zbiory zstępujące).

3.1.2. Segments in d directions

The same algorithm as before but in complexity $\mathcal{O}(d^k)$.

3.1.3. Segments in arbitrary direction

Theorem 3.1.1 ***(FPT for segment cover).*** *There exists an algorithm that given a family $\mathcal{P}$ of n segments (in any direction), a set of m points $\mathcal{C}$ and a parameter k, runs in time $f(k) \cdot (nm)^c$ for some computable function f and constant c, and outputs a subfamily $\mathcal{R} \subseteq \mathcal{P}$ such that $|\mathcal{R}| \leq k$ and $\mathcal{R}$ covers all points in $\mathcal{C}$.*

Proof. We will show such algorithm in FPT.

If there exist two segments a and b in $\mathcal{P}$, such that any point covered by a is also covered by b, then without loss of generality we can remove segment a from $\mathcal{P}$. We repeat this process until no such (a, b) pair exists.

Let us first assume that we reduced our instance to a kernel, where *any line* contains no more than k points.

Since any segment covers a set of colinear points, for such a kernel k segments can cover only at most k^2 points. Therefore, for the answer to be positive, the number of points has to be at most k^2. The number of segments is now bounded by k^4, since if we consider two *extreme* points covered by a given segment, then these pairs must be distinct, otherwise two segments would contain the same set of points. Since both the number of points and the number of segments is bounded by a function of k, this instance can be easily solved in time $O(f(k))$.

In remains to show how to construct the kernel.

Assume there exists a line l containing points $x_1, \dots x_t$, where $t \geq k+1$. Note that a segment that does not lie on l can cover only at most one of the points x_i. Therefore, out of points x_1, ..., x_{k+1}, at least one has to be covered by a segment that lies on l, let us fix x_i to be the first such point. Then, we can greedily choose a segment that lies on l, covers x_i, and also covers the largest number of points x_j for $j > i$.

Since we have at most $k+1$ choices to branch over and each choice adds a segment to the constructed solution, we obtain an algorithm with complexity $O(k^k)$.

3.2. APX-completeness for segments parallel to axis

3.2.1. Definition of MAX-(3,3)-SAT problem

Here we define MAXSAT problem.

Theorem 3.2.1 ***[Håstad, 2001]*** *Assume NP $\not\subseteq DTIME(2^{O(\log n \log \log n)})$. Then, there exists such constant $c > 0$, such for*

$$\epsilon' = \frac{c \log \log \log n}{\log \log n}$$

satifiable 3-SAT formulas cannot be distinguished from 3-SAT formulas where only $7/8 + \epsilon'$ of the clauses can be satisfied in polynomial time.

Lemma 3.2.1 *Given an instance of MAX-(3,3)-SAT with n variables and optimal result k, we can construct an instance of axis-parallel segments in 2D, which optimal result (even with 1/2-extension) is exactly $17n - k$.*

Theorem 3.2.2 ***(axis-parallel segment set cover with 1/2-extension is APX-hard).*** *For every $\epsilon > 0$, there doesn't exist an $(1+\epsilon)$-approximation scheme for unweighted geometric set cover with axis-parallel segments in 2D (even with 1/2-extension) (problem is APX-hard).*

Proof. Take any $1/(17 \cdot 8) > \epsilon > 0$. Take such n, that ϵ' from Theorem 3.2.1 is not greater than ϵ.

Let's assume that there exists an $(1+\epsilon)$-approximation scheme for unweighted geometric set cover with axis-pararell segments in 2D. We will construct an algorithm distinguishing instances of MAX-(3,3)-SAT in Theorem 3.2.1. Take two instances to be distinguished and using Lemma 3.2.1 let's construct two instances of geometric set cover, name the one constructed from satisfiable 3-SAT I_1 and the unsatisfiable 3-SAT as I_2.

Use $(1+\epsilon)$-approximation scheme for instances of geometric set cover, let's name the result of this approximation for an instance of problem I as $approx(I)$.

$$\frac{1}{8}n - \epsilon' > \frac{1}{8}n - \epsilon > 17n\epsilon - \epsilon > (16n+1)\epsilon - \epsilon = 16n\epsilon$$

$$16\frac{1}{8}n - \epsilon' > 16n + 16n\epsilon = 16n(1+\epsilon)$$

$$approx(I_2) \geq OPT(I_2) = 16\frac{1}{8}n - \epsilon' > 16n(1+\epsilon) = OPT(I_1)(1+\epsilon) \geq approx(I_1)$$

So we can distinguish these instances, since the satifiable instance will always yield lesser result in approximation scheme.

Therefore such approximation scheme cannot exist.

3.2.2. Reduction construction

Let's take some instance of MAX-(3,3)-SAT with variables $x_1, x_2 \dots x_n$ and clauses $C_1, C_2 \dots C_n$.

We will create gadgets for choosing the value of variables (*true* or *false*) and checking if the clauses are met (any of the variables were chosen).

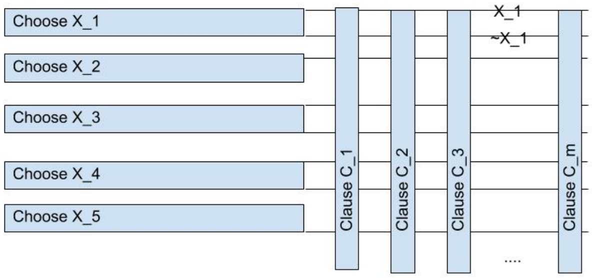


Figure 3.1: General scheme of reduction.

Choose x_i gadget

In Figure 3.2, we show a gadget that simulates a single variable x_i. It consists of six points A, B, C, D, E, F, and several segments. Selecting the segment marked with x_i to the solution will correspond to setting x_i to *true*, while selecting the segment marked with $\neg x_i$ to setting x_i to *false*. In the following lemmas, we show that this construction indeed models a binary variable.

First, note that in the gadget there are exactly two sets of three segments that cover all points A, B, C, D, E, F. These two sets of segments are marked in Figure 3.2 in blue and green, respectively.

Lemma 3.2.2 *Points A, B, C, D, E, F cannot be covered using less than 3 segments (even with 1/2-extensions).*

Proof. We need to take at least one segment on line ABC, because it's the only way to cover C. All other points (D, E, F) are not colinear, so we need at least 2 other segments to cover them.

Lemma 3.2.3 *If we choose both segments x_i and $\neg x_i$, we need to use at least 4 segments to cover all points A, B, C, D, E, F (even with 1/2-extensions).*

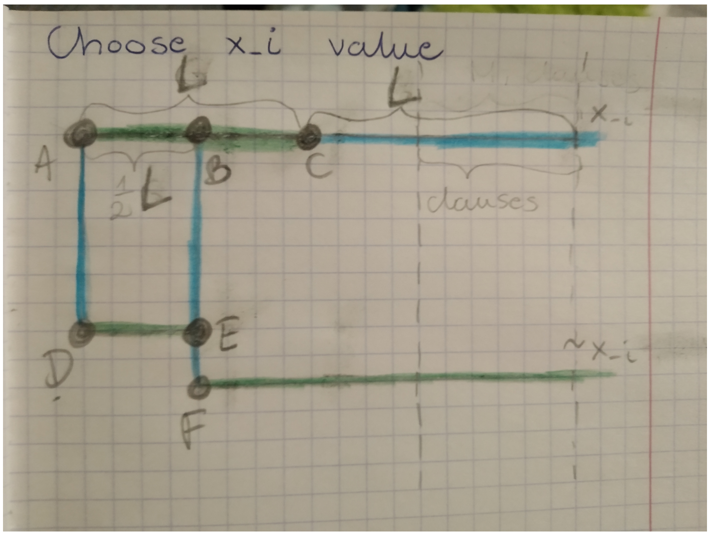


Figure 3.2: Scheme of choose x_i gadget.

Proof. Choosing both segments x_i and $\neg x_i$ we only cover points C (becuase B is too far away to be covered with 1/2-extension) and F.

The remaining points (A, B, D, E) are not colinear, so we need at least two more segments to cover them.

Lemma 3.2.4 *There exist a solution such that takes a segment x_i ($\neg x_i$) and 2 other segments, and covers all points A, B, C, D, E, F*

Proof. We can choose x_i, AD and BF.

Alternatively we can choose $\neg x_i$, AC, DE.

Robustness to 1/2-extension. Take a look at Figure 3.1. The points will be included in choose gadgets (horizontal boxes) and clause gadgets (vertical boxes).

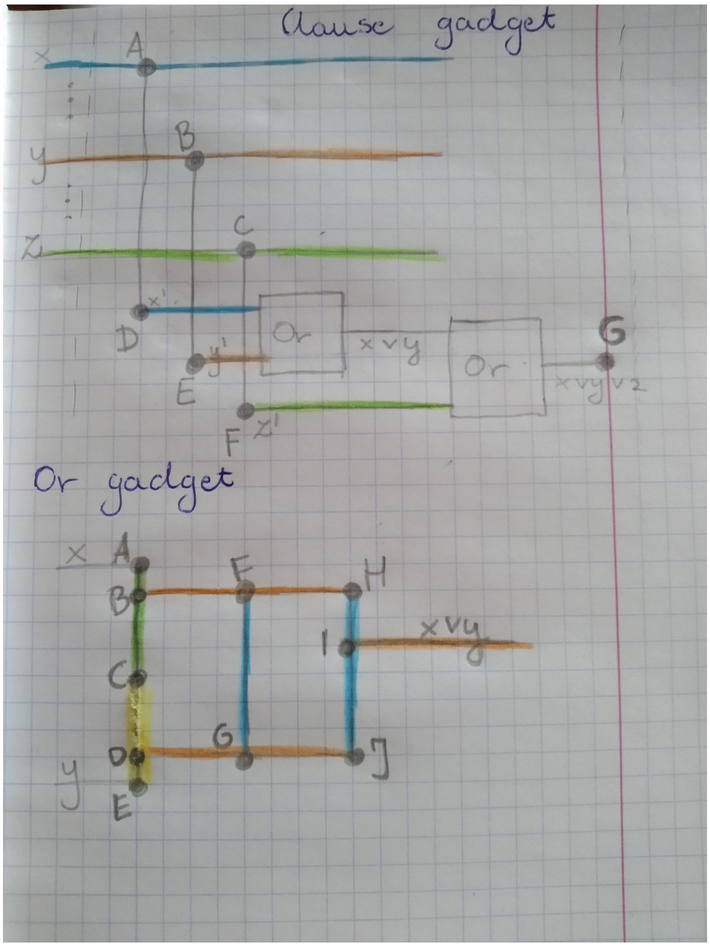
Since segment AC is very long and colinear with x_i, after 1/2-extension it will cover a significant part of segment x_i, even though x_i will not be chosen.

If we put all the clause gadgets in the area marked with **clauses** at gadget scheme in Figure 3.2, it is enough to prove that AC will not cover any points in the **clauses** area even with 1/2-extensions.

Lemma 3.2.5 *No points in **clauses** area can be covered by AC with 1/2-extension.*

Proof. Bear in mind that length of AC is equal to length of x_i. Area **clauses** takes a second half of the segment x_i and AC after extension will cover the first half of segment x_i.

Clause gadget



Lemma 3.2.6 *In order to cover D (E, F) point at least one of the segments AD (BE, CF) or x' (y', z').*

Lemma 3.2.7 *Points A and D can be covered with one additional segment x' only if x is already chosen. Otherwise they can be covered with one segment only by using AD.*

Lemma 3.2.8 *Points A, B, C, D, E, F can be covered with 3 segments and cannot be covered in less.*

Proof. There are 3 points that aren't pairwise connected, so they have to be covered with 3 separate segments.

Also there exists solution with 3 segments, ie. AD, BE, CF.

Or gadget

Lemma 3.2.9 *Points $A, B, C, D, E, F, G, H, I, J$ can be covered using at least 4 segments even with 1/2-extension.*

Lemma 3.2.10 *Points $A, B, C, D, E, F, G, H, I, J$ can be covered using 4 segments and segment $x \vee y$ can be chosen even with 1/2-extension only if at least one of the segments x or y is chosen.*

3.2.3. Proof that construction is sound

Lemma 3.2.11 *If there exists setting of values of variables that exactly k clauses are satisfied, we can cover all the points with $3n + 11m + (m - k)$ segments.*

3.2.4. Proofs of construction Lemma 3.2.1

Lemma 3.2.12 *Given an instance of MAX-(3,3)-SAT of size n with optimum solution k. For instance of geometric cover, constructed in the aforementioned manner, there exists a solution of weight $17n - k$.*

Proof. Let's name the assignments of the variables in MAX-(3,3)-SAT instance $y_1, y_2 \dots y_n$, Let's cover every clause with solution described in Lemma 3.2.4, in the i-th segment choosing the segment responsible for value y_i.

This solution uses $3n +$.

Lemma 3.2.13 *Given an instance of MAX-(3,3)-SAT of size n, and solution of size w to the instance of geometric cover, constructed in the aforementioned manner, there exists a solution to MAX-(3,3)-SAT of size at least $17n - w$.*

Proof. TODO

Proof of Lemma 3.2.1

Given an instance of MAX-(3,3)-SAT of size n with optimal result k. Let's construct an instance of geometric cover, constructed in aforementioned manner.

Given the Lemma 3.2.12, we know the optimal solution for the constructed geometric cover is at most $17n - k$ and since the k is optiomal solution for MAX-(3,3)-SAT, then according to Lemma 3.2.13 there doesn't exist a solution with cost lesser than $17n - k$.

3.3. Weighted segments

3.3.1. FPT for weighted segments with δ-extensions

Theorem 3.3.1 ***(FPT for weighted segment cover with δ-extensions)****. There exists an algorithm that given a family $\mathcal{P}$ of n weighted segments (in any direction), a set of m points $\mathcal{C}$ and a parameter k, runs in time $f(k) \cdot (nm)^c$ for some computable function f and constant c, and outputs a subfamily $\mathcal{R} \subseteq \mathcal{P}$ such that $|\mathcal{R}| \leq k$ and $\mathcal{R}^{+\delta}$ covers all points in $\mathcal{C}$.*

To solve this problem we will introduce kernel for slightly different problem: Weighted segment cover of points and segments. In shortcut: WSCPS.

Lemma 3.3.1 ***(Algorithm for kernel of WSCPS)****. There exists an algorithm that given a family $\mathcal{P}$ of n weighted segments (in any direction), a set of m_1 points $\mathcal{C}_1$ and m_2 segments $\mathcal{C}_2$ and a parameter k, runs in time $f(k) \cdot g(m_1, m_2) \cdot n^c$ for some computable functions f, g and constant c, and outputs a subfamily $sol \subseteq \mathcal{P}$ such that $|\mathcal{R}| \leq k$ and $\mathcal{R}$ covers all points in $\mathcal{C}_1$ and all segments in $\mathcal{C}_2$.*

Proof Only sketch for now.

We can compute dynamic programming $dp(A, B, z)$ – the best cost to cover at least whole segment A, B using at most z segments. A, B are all interesting points – ends of any segment given on the input or points given on the input. We can compute it in polynomial time.

Then we can create a new double weighted set (original weight, number of used segments from $\mathcal{P}$) – $\mathcal{P}_2$ that has only segments which never cover partially any segment from $\mathcal{C}_2$ (covers the whole segment or doesn't cover at all). In such $\mathcal{P}_2$ we can find solution $\mathcal{R}$ where any 2 segments have empty intersection (don't cover each other and don't meet at the ends). Because if we had such solution, we can merge these two segments and such segment there's also in $\mathcal{P}_2$.

In that case we can find kernel of $\mathcal{P}_2$ of size $k \cdot (m_1 + 2m_2)^2$, because we only need to take the best weight covering some subset of $\mathcal{C}_1 \cup \mathcal{C}_2$.

Lemma 3.3.2 ***Kernel in WSCPS***. *For segment cover, there is a kernel of size $f(k)$.*

Lemma 3.3.3 *If the biggest $(1/delta)^{k+1}$ spaces between points are filled, the whole segment is filled after δ-extensions of segments.*

3.3.2. W[1]-completeness for weighted segments in 3 directions

3.3.3. What is missing

We don't know FPT for axis-pararell segments without δ-extensions.

Chapter 4

Geometric Set Cover with lines

4.1. Lines parallel to one of the axis

When $\mathcal{R}$ consists only of lines parallel to one of the axis, the problem can be solved in polynomial time.

We create bipartial graph G with node for every line on the input split into sets: H – horizontal lines and V – vertical lines. If any two lines cover the same point from $\mathcal{C}$, then we add edge between them.

Of course there will be no edges between nodes inside H, because all of them are pararell and if they share one point, they are the same lines. Similar argument for V. So the graph is bipartial.

Now Geometric Set Cover can be solved with Vertex Cover on graph G. Since Vertex Cover (even in weighted setting) on bipartial graphs can be solved in polynomial time.

Short note for myself just to remember how to this in polynomial time:

Non-weighted setting - Konig theorem + max matching

Weighted setting - Min cut in graph of $\neg A$ or $\neg B$ (edges directed from V to H)

4.2. FPT for arbitrary lines

You can find this is Platypus book. We will show FPT kernel of size at most k^2.

(Maybe we need to reduce lines with one point/points with one line).

For every line if there is more than k points on it, you have to take it. At the end, if there is more than k^2 points, return NO. Otherwise there is no more than k^4 lines.

In weighted settings among the same lines with different weights you leave the cheapest one and use the same algorithm.

4.3. APX-completeness for arbitrary lines

We will show a reduction from Vertex Cover problem. Let's take an instance of the Vertex Cover problem for graph G. We will create a set of $|V(G)|$ pairwise non-pararell lines, such that no three of them share a common point.

Then for every edge in $(v, w) \in E(G)$ we put a point on crossing of lines for vertices v and w. They are not pararell, so there exists exactly one such point and any other line don't cover this point (any three of them don't cross in the same point).

Solution of Geometric Set Cover for this instance would yield a sound solution of Vertex Cover for graph G. For every point (edge) we need to choose at least one of lines (vertices) v or w to cover this point.

Vertex Cover for arbitrary graph is APX-complete, so this problem in also APX-complete.

4.4. 2-approximation for arbitrary lines

Vertex Cover has an easy 2-approximation algorithm, but here very many lines can cross through the same point, so we can do d-approximation, where d is the biggest number of lines crossing through the same point. So for set where any 3 lines don't cross in the same point it yields 2-approximation.

The problematic cases are where through all points cross at least k points and all lines have at least k points on them. It can be created by casting k-grid in k-D space on 2D space.

Greedy algorithm yields $\log|\mathcal{R}|$-approximation, but I have example for this for bipartial graph and reduction with taking all lines crossing through some point (if there are no more than k) would solve this case. So maybe it works.

Unfortunaly I haven't done this :(

I can link some papers telling it's hard to do.

4.5. Connection with general set cover

Problem with finite set of lines with more dimensions is equivalent to problem in 2D, because we can project lines on the plane which is not perpendicular to any plane created by pairs of (point from $\mathcal{C}$, line from $\mathcal{P}$).

Of course every two lines have at most one common point, so is every family of sets that have at most one point in common equivalent to some geometric set cover with lines?

No, because of Desargues's theorem. Have to write down exactly what configuration is banned.

Chapter 5

Geometric Set Cover with polygons

5.1. State of the art

Covering points with weighted discs admits PTAS [Li and Jin, 2015] and with fat polygons with δ-extensions with unit weights admits EPTAS [Har-Peled and Lee, 2009].

Although with thin objects, even if we allow δ-expansion, the Set Cover with rectangles is APX-complete (for $\delta = 1/2$), it follows from APX-completeness for segments with δ-expansion in Section 3.2.

Covering points with squares is W[1]-hard [Marx, 2005]. It can be proven that assuming $SETH$, there is no $f(k) \cdot (|\mathcal{C}| + |\mathcal{P}|)^{k-\epsilon}$ time algorithm for any computable function f and $\epsilon > 0$ that decides if there are k polygons in $\mathcal{P}$ that together cover $\mathcal{C}$, *Theorem 1.9* in [Marx and Pilipczuk, 2015].

Chapter 6

Conclusions

Bibliography

[Har-Peled and Lee, 2009] Har-Peled, S. and Lee, M. (2009). Weighted geometric set cover problems revisited. *Journal of Computational Geometry*, 3.

[Håstad, 2001] Håstad, J. (2001). Some optimal inapproximability results. *J. ACM*, 48(4):798–859.

[Li and Jin, 2015] Li, J. and Jin, Y. (2015). A PTAS for the weighted unit disk cover problem. *CoRR*, abs/1502.04918.

[Marx, 2005] Marx, D. (2005). Efficient approximation schemes for geometric problems? In Brodal, G. S. and Leonardi, S., editors, *Algorithms – ESA 2005*, pages 448–459, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Marx and Pilipczuk, 2015] Marx, D. and Pilipczuk, M. (2015). Optimal parameterized algorithms for planar facility location problems using voronoi diagrams. *CoRR*, abs/1504.05476.