Chapter 3

The Fréchet Metric

3.1 Motivation

As already mentioned in section 2.2 on page 5, the Hausdorff metric lacks in using the additional information which comes with the parameterization of the given objects, provided that there is a parameterization. In this chapter, however, we assume the geometric objects to be given by some parameterization.

If we study those objects like curves it turns out that a different metric called Fréchet metric is more appropriate. It was first described in [Fré06]. Especially in the calculus of variations this is the standard metric considered, see [Ewi85] for example. The Fréchet metric is also sometimes simply more usefull than the Hausdorff metric especially in solving problems in pattern recognition as you might see in [WN94].

The Fréchet metric is also the only metric for curves for which there is an algorithm which simplifies a piecewise affine curve by searching for a piecewise affine curve with the least possible number of pieces within a neighborhood of the original curve where neighborhood is defined with respect to that metric (see [GHMS93]). This may be surprising but we think it reflects only fact that a global optimum with respect to, say the Hausdorff metric would be as difficult as it does not make sense because, as mentioned before, the Hausdorff metric does not necessarily expresses what we want. Hence existing approximation software makes some compromise between local and global features. Thus it is not easy to describe what they do in terms of a metric.

So let us try to give it more formal definitions.

3.2 Definitions

Let (X, δ_X) be a fixed metric space and $d \in \mathbb{N}$ a fixed constant. This will be the intrinsic dimension of the geometric objects we consider.

The geometric objects we will consider in this chapter —let us call them ***objects*** for short— are continuous mappings $f : A \to X$ where $A \subseteq \mathbb{R}^d$ is homeomorphic to $[0, 1]^d$. Sometimes more complicated domains are considered which

need not to be mutually homeomorphic. But this should not be the main problem here.

We would like to introduce the abbreviation $\boldsymbol{\sigma : A \xrightarrow{\sim} B}$ for the fact that $\sigma\colon A \to B$ is an orientation preserving homeomorphism.

Usually objects $f\colon A \to X$ and $g\colon B \to X$ are identified if there is a $\sigma\colon A \xrightarrow{\sim} B$ with $f = g \circ \sigma$. In this case we say f is achieved from g by a (orientation preserving) reparametrization. It is easy to compare functions with the same domain just by evaluating the pointwise minimum. Unfortunately the result of this comparison depends heavily on the parameterization. Thus the idea behind the definition of the Fréchet metric is to choose parameterizations among all valid reparametrizations of an object which match best.

For two objects $f\colon A \to X$ and $g\colon B \to X$ the ***Fréchet distance*** is defined by

$$\delta_F(f,g) := \inf_{\sigma: A \xrightarrow{\sim} B} \sup_{x \in A} \delta_X\big(f(x), g(\sigma(x))\big)$$

or $+\infty$ if no such σ exists, which would be only the case if we allowed the domains of our objects to be more general. This distance function is a pseudo metric*.

Note: Defining $f \sim g :\Leftrightarrow \delta_F(f,g) = 0$ and considering the equivalence classes $f_{/\sim} := \{\widetilde{f}\colon \widetilde{f} \sim f\}$ as the *real* objects makes $\delta_F(f_{/\sim}, g_{/\sim}) := \delta_F(f,g)$ to a metric on these "real objects". • For $d = 1$ these equivalence classes are called ***oriented Fréchet curves*** and •• for $d = 2$ they are called ***oriented Fréchet surfaces***.

From the algorithmic point of view these objects are too general. Therefore we want to define an object $f\colon A \to X$ to be ***simplicial*** if and only if A is the underlying space of a finite simplicial complex of dimension d and f is affine on each simplex of this complex. See section 2.3 on page 6 for a definition of these terms. Then the ***size*** of the object is the complexity of the complex, i.e. the number of simplices in it. • For $d = 1$ this means that A should be subdivided into finitely many subintervals and that f should be affine on each subinterval. This means that f should describe a polygonal chain. •• For $d = 2$ this means that A should be triangulated into finitely many triangles and that f should be affine on each triangle.

A simplicial object $f\colon A \to X$ can be described by a finite structure describing the simplicial complex and by the values of f on the finitely many corners of the simplices in the complex. For example for $d = 2$ it suffices to specify the structure of the oriented triangulation of A and to specify what f does on the corners of the triangles.

Furthermore, and for the sake of simplicity, let us assume that X is a finite dimensional euclidian vector space like $\mathbb{R}^3$ or so. Finally let us assume that all corners of the simplices we consider as input for our algorithms have rational coordinates.

---

*That means $\delta_F(f,f) = 0$
$\delta_F(f,g) = \delta_F(g,f)$
$\delta_F(f,h) \leqslant \delta_F(f,g) + \delta_F(g,h) \quad \forall f,g,h.$
See [Ewi85] for the proofs. But it is not really complicated.

Now we can address the two problems considered in this context:

▷ Given simplicial objects f and g of size n and m. What is $\delta_F(f,g)$? This is called the ***computation*** of the distance.
▷ Given simplicial objects f and g of size n and m. Given furthermore a real ε. Does $\delta_F(f,g) \leqslant \varepsilon$ hold? This easier problem is called the ***decision problem*** for the distance.

3.3 Previous work

For $d = 1$ the decision problem can be solved in $\mathcal{O}(nm)$ time and the distance can be computed in $\mathcal{O}(nm \log nm)$ time [AG95]. For $d > 1$ nothing was known. The algorithms for $d = 1$ made essentially use of the linear ordering in $\mathbb{R}^1$. The lack of ordering in higher dimensions seems to make it difficult to apply the same ideas here.

3.4 The main theorem

The decision problem for the Fréchet metric for simplicial objects with intrinsic dimension $d \geqslant 2$ is NP-hard even if $X = \mathbb{R}^2$ and one object describes a single fixed triangle.

3.5 Background

Cook [Coo71] proved the NP-completeness of 3SAT by directly showing that every problem in NP can be reduced in polynomial time onto 3SAT. With this first NP-complete problem other problems could be proved to be NP-complete simply by reducing known NP-complete problems to them.

3.5.1 Definitions

▷ A boolean formula is said to be in ***conjunctive normal form*** if and only if it is a conjunction of disjunctions of literals. A ***literal*** is a variable or a negated variable. The disjunctions are also called ***clauses***.
▷ A boolean formula is said to be in ***3-conjunctive normal form*** if and only if it is in conjunctive normal form and every clause consists of exactly three literals, where we allow one variable to appear more than once in a clause.
▷ A boolean formula is said to be in ***3,4-conjunctive normal form*** if and only if it is in 3-conjunctive normal form and every variable occurs at most four times, where repeated occurrences of one variable in one clause are counted repeatedly.
▷ A boolean formula φ in conjunctive normal form is said to be ***planar*** if and only if the bipartite graph B_φ is planar, where the vertices of B_φ are the variables and clauses of φ and the edges are exactly the pairs (v, c) for which v is a variable occurring in the clause c. This definition differs slightly from the one given in

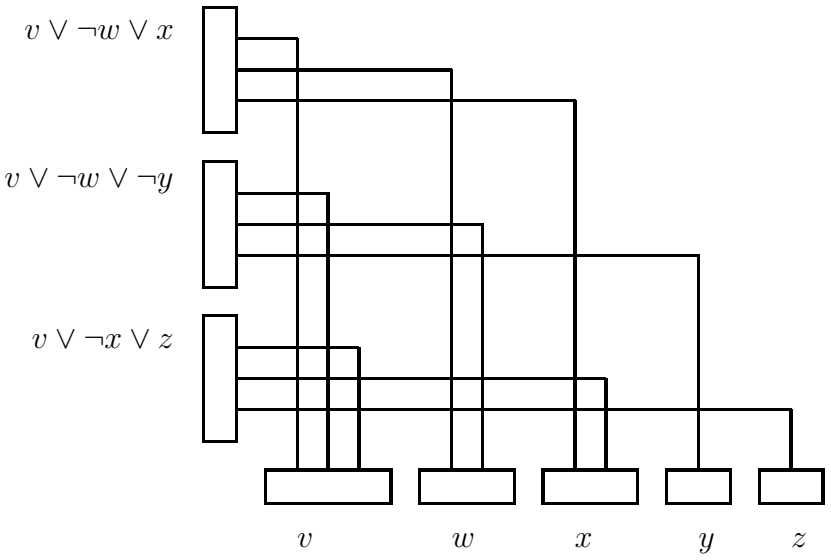
[Lic82], but it is sufficient to know that a planar formula in the sense of [Lic82] is planar in the definition above, too.
▷ Let *yellow submarine*-SAT denote now the problem of deciding whether a given boolean formula is satisfiable, which is in *submarine*-conjunctive normal form and furthermore is *yellow*. Of course, these words *yellow* and *submarine* are only placeholders to illustrate this way of speaking.

3.5.2 Known reductions

In [Tov84] it was shown that 3,4-SAT is NP-complete. The idea used to show one variable has to occur no more than four times is indeed very simple; we can replace each variable x occurring k times by $x_0, ..., x_{k-1}$ and add the clauses $(x_i \vee \neg x_{i+1 \bmod k})$. If we write $(x_i \vee \neg x_{i+1 \bmod k})$ as $(x_i \vee x_i \vee \neg x_{i+1 \bmod k})$ this gives a clause consisting of exactly three literals and every x_i occurs exactly four times.* The only reason to mention this is to notice that this process keeps the planarity of the formula.

In [Lic82] is shown that planar 3-SAT is NP-complete. With the idea in [Tov84] it can be shown that planar 3,4-SAT is also NP-complete. A careful reading of the construction in [Lic82] used to show planar 3-SAT to be NP-complete makes it clear that for a given formula of length n in 3,4-conjunctive normal form not only an equivalent instance of planar 3,4-SAT consisting of some formula φ can be computed in in n polynomial time, but also an embedding of B_φ can be computed in a very simple way.

To understand this we briefly repeat the construction in [Lic82]. First all variables are placed on a horizontal line and all clauses are placed on a vertical line. Then the connections are made in an obvious manner. Take for example the formula $(v \vee \neg x \vee z) \wedge (v \vee \neg w \vee \neg y) \wedge (v \vee \neg w \vee x)$. Then the straight forward embedding is shown below.



*This is a dirty trick. Strictly speaking in [Tov84] is actually shown how to avoid multiple occurrences of one variable within one clause. But this is not important for our proof.

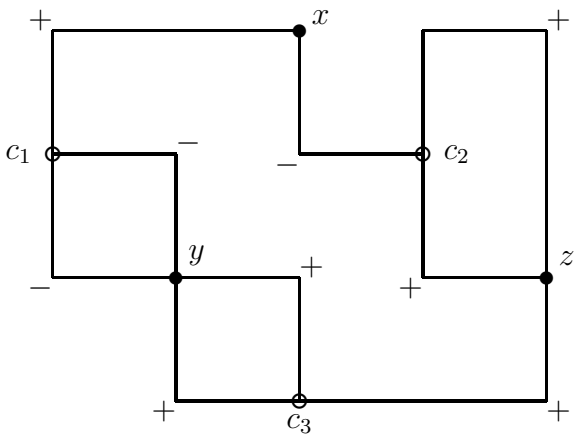
So far there is obviously no problem in computing the embedding. But of course, there will be crossings between the edges. These crossings are eliminated step by step by substituting each crossing by a certain planar subgraph which will introduce some new variables and clauses (just at the place where the crossing was) such that the new formula is satisfiable if and only if the old formula is satisfiable. This construction can be done in quadratic time.

We can furthermore* assume that B_φ is embedded on a $m \times m$-grid where the vertices are non adjacent grid points and the edges are grid point disjoint paths on the grid lines and m is linear in n.

3.5.3 Straight forward extensions

Altogether the following problem —let us call it ***grid3sat***— is also NP-complete. We are given a $m \times m$-grid, in which some non adjacent grid points are distinguished as clauses and some as variables. The variables are connected with the clauses by vertex disjoint paths on the grid. We associate a sign with every path indicating whether the corresponding variable is negated in the corresponding clause or not. The question is, whether the formula described in this funny way is satisfiable or not.

The reductions of grid3sat to some other problems will be computationally trivial. Take for example the following instance of grid3sat.

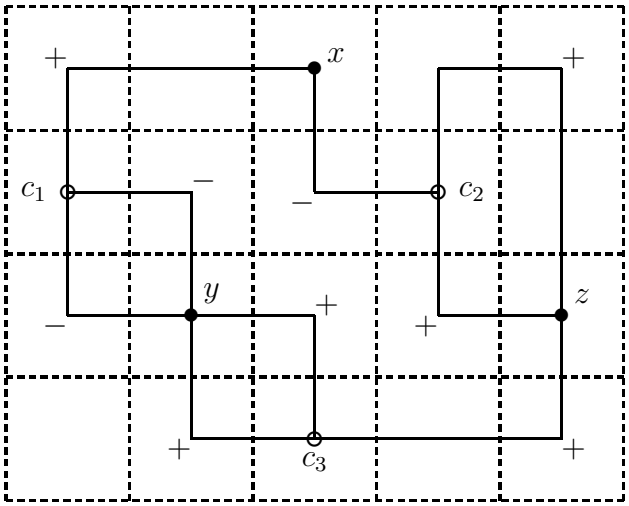


The filled circles indicate variables, the empty ones indicate clauses. The formula described by the drawing above is

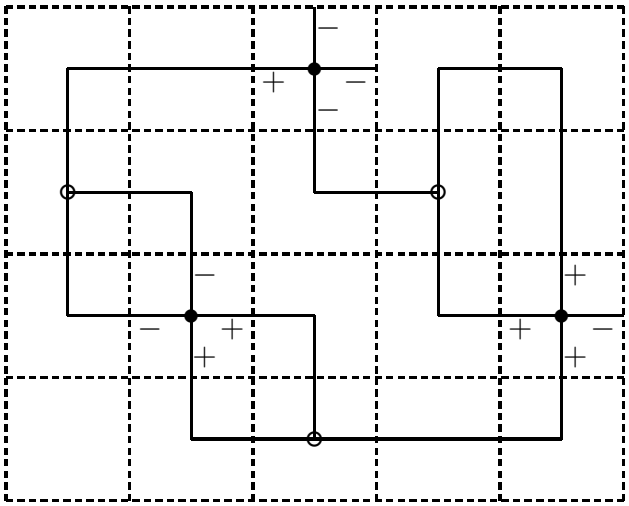
$$(x \vee \neg y \vee \neg y) \wedge (\neg x \vee z \vee z) \wedge (y \vee y \vee z).$$

All the reductions we will consider later on from grid3sat to some other problem begin as follows. First the grid is subdivided into quadratic pieces centered at the grid points subsequently called ***components***.

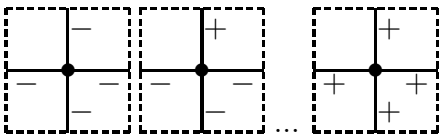
*This is not mentioned in [Lic82] but should be clear.



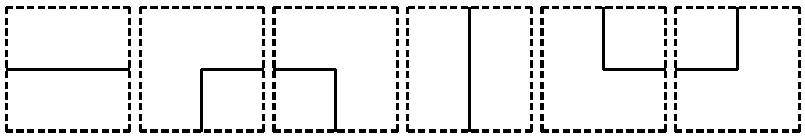
In order to reduce the number of different components we assume each component containing a variable (subsequently called ***variable component***) to have exactly four exits.* Furthermore we draw the signs on the paths directly on the variables.



Thus we have 26 different possible components. These are $2^4 = 16$ different variable components,

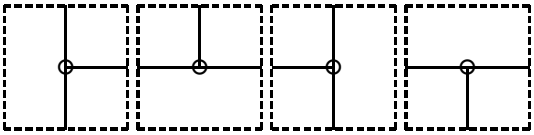


6 different so called ***connection components***,



*This will not cause any problems since there are no adjacent variables.

and 4 different so called ***clause components***.



In the reductions we will have some problem specific constructions for each of the different possible components. The reductions will be computationally trivial since we only have to insert these constructions everywhere the components appear on the grid.

Historical note. Having tried to disprove the main problem in this chapter, i.e. having tried to give a polynomial time algorithm for the decision problem for the Fréchet metric we have tried first to solve a problem which seemed to be easier. But even that turned out to be NP-hard. See chapter 5 on page 61 for details. This was the reason for us to try to prove the decision problem to be NP-hard.

3.6 The selection problem

In this section we discuss a more or less artificial problem we would like to call the selection problem. We will show this problem to be NP-hard. In the next sections we will reduce the decision problem for the Fréchet metric to the selection problem.

Remark. In this and the following sections of this chapter we will introduce a bunch of ridiculous funny looking numeric constants such as 0.000000016, 0.0000009, 0.000000049, 0.00000052, 0.0054, 0.99999999, 1.09 or 3.0000001. These constants seemed to be randomly chosen and therefore somehow arbitrary. In fact this is not true.

There is some clearance in choosing those numbers. That means the proofs would probably still work even if we would choose numbers which differ from the ones given here a little bit. So much is true. But We have the impression that there is only very little clearance in some of these numbers. Especially the constant 3.0000001 which you will encounter in section 3.11 seems to look as if it would have to be an arbitrary number > 3. It is not known whether the statement which is proven in that section is true for all numbers > 3 but it is clear that the proof would not work for arbitrary numbers. So what we basically did is to try to find a reasonable large value for it for which the proof works. The fact that this value is so close to 3 only reflects the tightness in the constructions.

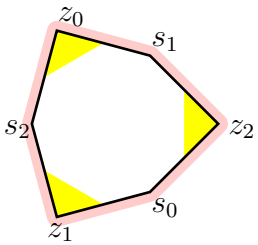
3.6.1 Problem definition

A ***symbol placement*** is a quintuple (x, y, c, s, n) where x, y, c, s denote rational numbers with $c^2 + s^2 = 1$ and $n \in \{2, 3\}$.

To each symbol placement we associate a symbol which is a hexagon* in the euclidian plane for which some vertices are distinguished to be ***yellow***, some are not. Here is the symbol for the symbol placement $(0, 0, 1, 0, 3)$.

*considered as a convex body, not as a boundary of a convex body

Symbol figure Symbol coordinates

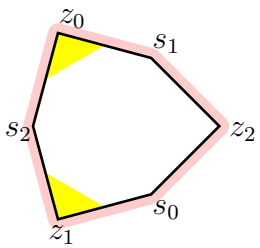


z_0=(-.5450000, .9439677)
z_1=(-.5450000,-.9439677)
z_2=(1.0900000, 0)
s_0=(.3989677,-.6910323)
s_1=(.3989677, .6910323)
s_2=(-.7979354, 0)

The pink region around the hexagon indicates the $\frac{\nu}{2}$ neighborhood of the hexagon, where $\boldsymbol{\nu} := 0.2$.

We will explain the geometry behind these fancy coordinates in subsection 3.6.2 on page 27. For the moment they are just the coordinates used in the subsequent drawings.* In a symbol placement (x, y, c, s, n) the entries x, y represent a translation and the entries c, s represent a rotation of the points above.** A point (p, q) for a symbol placement $(0, 0, 1, 0, n)$ corresponds to a point $(x+pc-qs, y+ps+qc)$ for the symbol placement (x, y, c, s, n).

There are two types of symbols. A symmetric one, associated with a symbol placement of the form $(x, y, c, s, 3)$ and an asymmetric one, associated with a symbol placement of the form $(x, y, c, s, 2)$. The asymmetric version for $(0, 0, 1, 0, 2)$ will be visualized as follows



The only difference is the coloring of the symbol. In the symmetric symbol the vertices z_0, z_1, z_2 are said to be yellow and in the asymmetric symbol only the vertices z_0, z_1 are said to be yellow. So the last number of the quintuple is actually a notation for the number of yellow vertices of the symbol.

Further definitions. For any $\mu > 0$ and two symbol edges (s, z) and (s', z') which are in general symbol edges from two different symbols we say that (s, z) and (s', z') are ***μ-close*** if and only if z and z' are both yellow and furthermore $||s - s'||, ||z - z'|| \leqslant \mu$ holds.

*These drawings are made by hand coded postscript. Thus the geometric calculations needed for the drawings are basically done by the postscript interpreter with the only exception that the coordinates printed here and elsewhere explicitly came from a separate file and are calculated by a short GNU bc program.

**Since c and s have to be rational, not all rotations are allowed. But this is far less restrictive as one can think. In [CDR92] you may also read how you can calculate approximations for arbitrary angles fast.

An ***ensemble*** is a set of symbol placements such that for any two symbols (S, S') the following conditions hold:

1. They do not intersect.

2. If there exist $x \in S$ and $x' \in S'$ with $||x - y|| < \nu$ then there exist edges (s_j, z_i) of S and $(s'_{j'}, z'_{i'})$ of S' which are μ-close to each other where here and in the following we will define $\boldsymbol{\mu} := 0.0000009$. In the following we will say in this case that z_i and $z'_{i'}$ are ***close*** to each other.

A ***selection*** of an ensemble is a subset of the yellow vertices of the symbols. A selection is said to be ***complete*** if and only if for every symbol at least one of its yellow vertices is in the selection. A selection is said to be ***feasible*** if and only if no two of its vertices are close to each other.

The ***selection problem*** is the following problem. Instance: Ensemble of n Symbols. Question: Does there exists a complete and feasible selection?

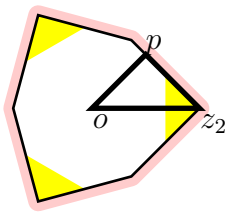
3.6.2 The geometry of the symbol

To understand the properties of the given coordinates for the symbols it is necessary to distinguish between the following notions of the symbol: the "ideal" symbol, the "real" symbol and the "scaled" symbol.

The ***ideal symbol*** for the symbol placement (x, y, c, s, n) is a hexagon with endpoints ${}^{\text{ideal}}z_0, {}^{\text{ideal}}s_2, {}^{\text{ideal}}z_1, {}^{\text{ideal}}s_0, {}^{\text{ideal}}z_2, {}^{\text{ideal}}s_1$ in counterclockwise direction and uniquely defined by following properties.

- All edges of the hexagon are of the same length.
- ${}^{\text{ideal}}s_0, {}^{\text{ideal}}s_1, {}^{\text{ideal}}s_2$ as well as ${}^{\text{ideal}}z_0, {}^{\text{ideal}}z_1, {}^{\text{ideal}}z_2$ form each a regular triangle with barycentre at (x, y) which will be subsequently called the ***center of the symbol*** as well.
- The angles at ${}^{\text{ideal}}z_0, {}^{\text{ideal}}z_1, {}^{\text{ideal}}z_2$ are right.
- To be more precise we demand that $z_2 = (x + 1.09c, y + 1.09s)$.

From the definition follows that a disk of radius 1.09 centered at (x, y) touches z_0, z_1, z_2 on its boundary and contains s_0, s_1, s_2 in its interior. Furthermore the distance of the center of a symbol to any of its edges is the same as to any line going though its edges and is equal to $1.09\frac{1}{2}\sqrt{2}$.



Proof. Let p be the point on the line $z_2 s_1$ nearest to o. Then $\angle p z_2 o$ is right. Since $\angle z_2 s_0 s_1$ is right the angle $\angle z_2 o p$ is $= \frac{1}{4}\pi$ as well as $\angle o z_2 p$. Thus we have $\delta(o, p) = \delta(o, z_2)\frac{1}{2}\sqrt{2} = 1.09\frac{1}{2}\sqrt{2}$. And, as you might see, p lies actually *on* the segment $z_2 s_1$. □

The ***real symbol*** for the symbol placement $(0, 0, 1, 0, n)$ is what we actually have seen so far and it is basically the same as the corresponding ideal symbol with the exception that the coordinates of the points are rounded to the nearest decimal description with 7 digits after the decimal point. This means that the real coordinates are rational and differ from the ideal coordinates by at most $\frac{1}{2} \cdot 10^{-7}$ which leads to a distortion of at most $\varrho := \frac{1}{2}\sqrt{2} \cdot 10^{-7} < 10^{-7}$.

The ***scaled symbol*** is the ideal symbol scaled around its barycentre by a factor of $1 + \alpha$ where $\alpha := 3 \cdot 10^{-7}$.

Claim. Let S, S' be scaled symbols. Let I be an isometric mapping which distorts the points of S' by a distance of at most $\delta := 2 \cdot 10^{-7}$ such that S and $I(S')$ have exactly one edge e in common. Furthermore this edge e of S has to be δ-close to $e' := I^{-1}(e)$ considered as an edge of S'. Let $\widetilde{S}, \widetilde{S}'$ be the corresponding real symbols as well as $\widetilde{e}$ and $\widetilde{e}'$ the corresponding edges on them. *Then* **(a)** the convex hulls of $\widetilde{S}, \widetilde{S}'$ do not intersect and **(b)** $\widetilde{e}$ and $\widetilde{e}'$ are μ-close together.

Proof.

(a)

Let $\overline{S}, \overline{S}'$ denote the convex hulls of the corresponding ideal symbols. Then we have $\delta_H(\overline{S}, I(\overline{S}')) = 1.09\alpha\sqrt{2}$ hence $\delta_H(\overline{S}, \overline{S}') \geqslant 1.09\alpha\sqrt{2} - \delta = 2.624... \cdot 10^{-7}$.

Assume now that the convex hulls of $\widetilde{S}$ and $\widetilde{S}'$ would intersect. That means that there must be an intersection point, let us call it p. By $\delta_H(p, \overline{S}), \delta_H(p, \overline{S}') \leqslant \varrho$ it would hold $\delta_H(\overline{S}, \overline{S}') \leqslant 2\varrho = \sqrt{2} \cdot 10^{-7} < 2.624... \cdot 10^{-7}$.

(b)

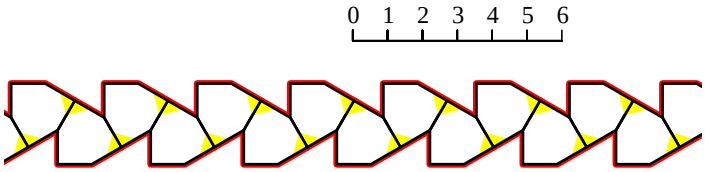
Let p denote a vertex of e and let p' denote the corresponding vertex of e'. Analogously let $\widetilde{p}, \widetilde{p}'$ the corresponding vertices of $\widetilde{e}, \widetilde{e}'$. Then it remains to show that $||\widetilde{p} - \widetilde{p}'|| \leqslant \mu$ holds.

Finally let $\overline{p}, \overline{p}'$ denote the corresponding vertices of the ideal symbols. Since on ideal symbols every vertex is at most 1.09 far apart from its center the distance between the corresponding vertices of the ideal and the scaled symbol is $\leqslant 1.09\alpha$. Now we have $||\widetilde{p}-\widetilde{p}'|| \leqslant ||\widetilde{p}-\overline{p}||+||\overline{p}-p||+||p-p'||+||p'-\overline{p}'||+||\overline{p}'-\widetilde{p}|| \leqslant \varrho+1.09\alpha+\delta+1.09\alpha+\varrho = 2\varrho+2.18\alpha+\delta = (\sqrt{2}+7.54)10^{-7} = 8.954... \cdot 10^{-7} \leqslant 9 \cdot 10^{-7} = \mu$. □

3.6.3 The selection problem is NP-hard

We reduce grid3sat to the selection problem. Let us remember. We are given an instance of grid3sat. We want to construct an appropriate instance of the selection problem. In order to do this we first scale the given grid such that the distance between two grid points is 180 000 000. Then we build each component with a finite number of symbol placements by a plan which is fixed for each of the 26 component types. Not really. We have to use two different versions for each of the 6 different connection components. This is because there is an implicit direction in each path (i.e. from the variable towards the clause) which we have to take into account when actually building these components. Therefore we actually use a stock of 32 different components.

Let us first look to the most simple form of a connection component. In the following picture we will show one which is directed from the left to the right.

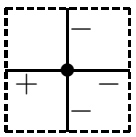


If not stated otherwise all symbols mentioned here are asymmetric ones. The symbols shown in this and the following pictures in this subsection are meant to be scaled symbols with edges matching exactly. In order to get a proper ensemble we have to use symbol placements describing real symbols. Due to the fact that we have to use rational numbers to describe these symbol placements there will be some quantization error. Thus for each symbol S we choose a symbol placement such that there is an isometric mapping I which maps S onto the scaled symbol described by the symbol placement such that all points in the 42-neighborhood of S are mapped onto points which are $\frac{1}{4}\delta$-close to the original points. The set of symbol placements we get that way will be an ensemble (use claim (*) for a proof).

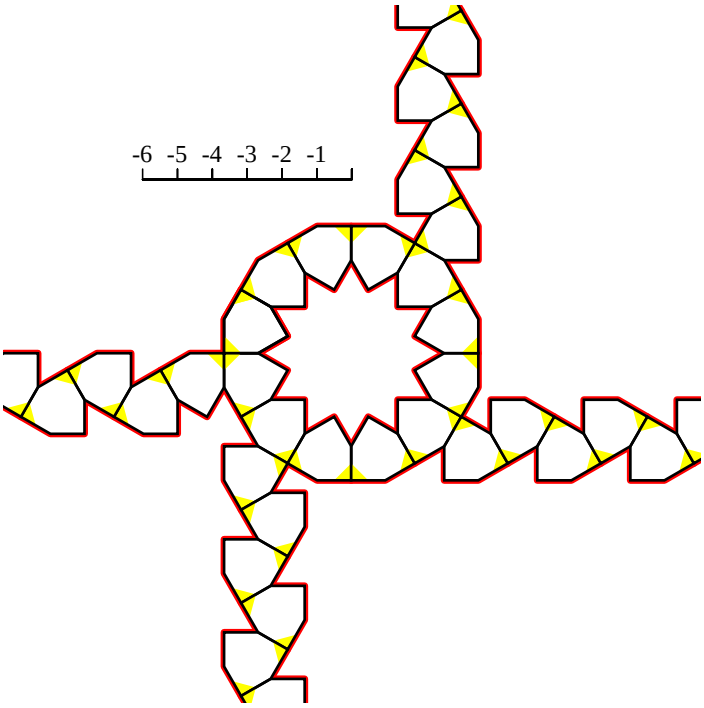
In the chain there is a pattern consisting of two symbols repeating itself. Each pattern increases the length of a value which is clearly < 3. Thus we can achieve any desired length > 3 within an error of $\leqslant 1.5$. For a chain of a length of approximately 90 000 000 we have to use more than 30 000 000 repeated patterns containing more than 60 000 000 symbols in total. If we have such a chain of symbols $g_0, ..., g_{n-1}$ with $n > 60\,000\,000$ and we are given a vector $\vec{v}$ with $||\vec{v}|| \leqslant 6$ then we are able to translate the symbol g_i by $\frac{i}{n}\vec{v}$ and the corresponding ensemble obtained in the way described above still remains to be an ensemble. For a proof note that two adjacent symbols are shifted against each other by $\frac{1}{n}\vec{v}$ and that $||\frac{1}{n}\vec{v}|| \leqslant \frac{6}{n} \leqslant 10^{-7} = \frac{1}{2}\delta$ and use claim (*).

Now we want to characterize the possible selections of an ensemble for such a chain. Imagine these selections as chains of dominoes. Each domino corresponds to a symbol. A domino is said to be fallen to the left if and only if the leftmost yellow vertex of the symbol is chosen to be in the selection. To be fallen to the right is defined in an analogous manner. If the selection is complete every domino is fallen. The selection to be feasible means that dominoes are fallen in a way dominoes usually fall, that is not towards each other. Altogether in a feasible and complete selection all dominoes are fallen either all in the same direction or from a certain point all outwards.

Now we are ready to discuss the variable component. Take for example this one.



The following picture shows the central part of the appropriate component.

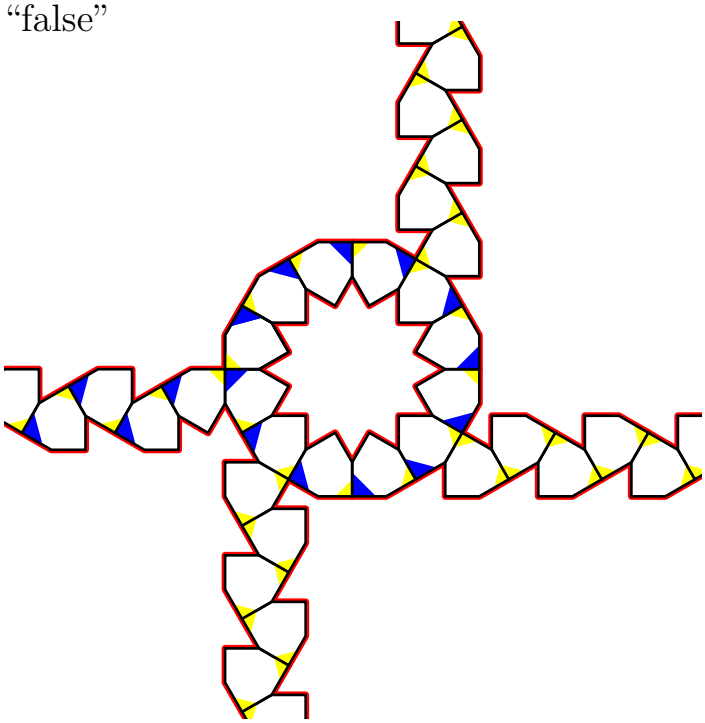


It consists mainly of an inner cycle and four chains which are all oriented outwards. The chains are long enough to able to be shifted in the way described above in order to make them fit to the other components.

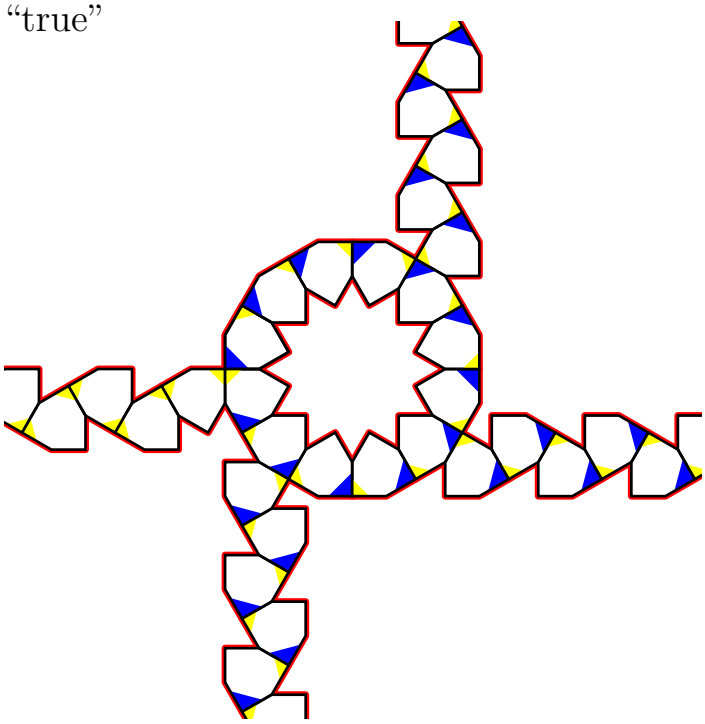
The variable component works as follows. In some sense we can understand the inner cycle in the same way we understood a chain of symbols, namely as a chain of dominoes. Do you remember? In a complete and feasible selection all dominoes are fallen either all in the same direction or from a certain point all outwards. The only difference is that in a cycle the third possibility does not exist. Thus all dominoes has to be fallen in the same direction and this means here either all in clockwise direction or all in counter clockwise direction. In the following we will refer to the first case as “false” and to the latter case as “true”.

In the following pictures those vertices are highlighted for which the selected vertex is determined by this.

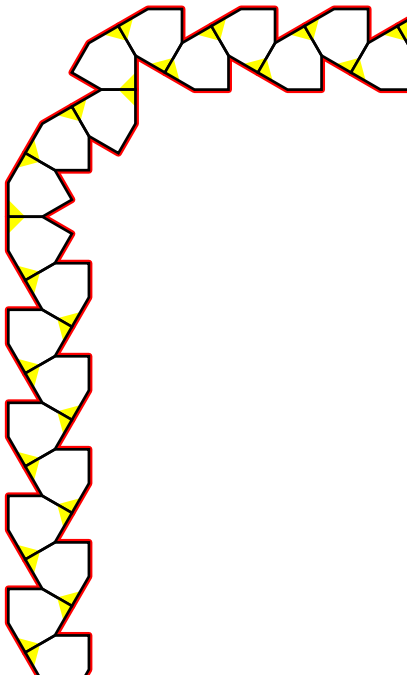
“false”



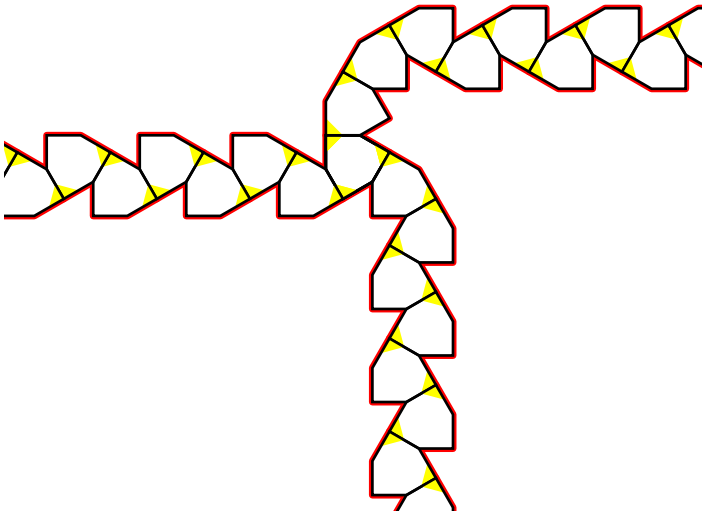
“true”



And now for the connection components. The straight one we have already seen. A kinky one would look like the following one.



Finally the clause components contain exactly one symmetric symbol which is exactly in the center and some chains of asymmetric symbols around, as you may see in the following picture.



In a complete and feasible selection there has to be exactly one yellow vertex of the central symmetric symbol in the selection, and this enforces the dominoes of one of the chains to be fallen outward thereby enforcing a certain state of the variable connected by this chain. Thus a complete and feasible selection can only exist if there is a satisfying truth assignment for the given instance of grid3sat.

On the other hand, if there exists a satisfying truth assignment for the given instance of grid3sat then we are able to select vertices in the inner cycles of the

variable components in a way described above and we are able to select one satisfying literal in each clause. If we take the corresponding vertex of the symmetric symbol into the selection and if we complement the selection in a way that the dominoes fall towards the central symbol of a clause if they represent a path to a literal which do not satisfy the clause and towards the inner cycle of the variable otherwise then this will be a complete and feasible selection.

3.7 Back to the Fréchet metric

Next we prove the decision problem for the Fréchet metric to be NP-hard. We do this by reducing the problem to the selection problem. This we will do in several steps. In the following section we discuss the basic tool in the reduction which will be called the ***gadget***. In section 3.9 we will define the reduction and in sections 3.10 and 3.11 we will explain how the reduction works.

Some oddities in the proof are due to the fact that we want to strengthen the result afterwards a little bit, see section 3.12 on page 55 for details.

3.8 The gadget

For each symbol placement we define a so called ***gadget***. A gadget consists of a labeled plane* graph consisting of nine vertices $x_0, x_1, x_2, y_0, y_1, y_2, z_0, z_1, z_2$. Each vertex v is assigned to a closed disk $\hat{v}^3$ of radius 3 centered at a point $\overline{v}$.

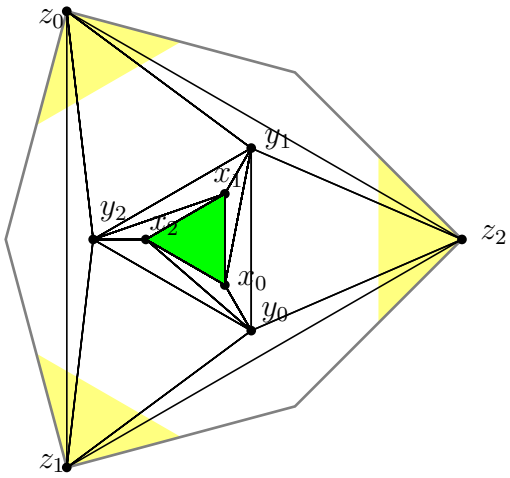
There are two types of gadgets, namely ***symmetric gadgets*** and ***asymmetric gadgets***. A symmetric gadget is a gadget for a symbol placement of the form $(x, y, c, s, 3)$ whereas an asymmetric gadget is a gadget for a symbol placement of the form $(x, y, c, s, 2)$.

The points $x_0, x_1, x_2, y_0, y_1, y_2, z_0, z_1, z_2$ as well as $\overline{x}_0, \overline{x}_1, \overline{x}_2, \overline{y}_0, \overline{y}_1, \overline{y}_2, \overline{z}_0, \overline{z}_1, \overline{z}_2$ for a symbol placement (x, y, c, s, n) are defined in quite in the same way as mentioned in subsection 3.6.1 on page 25, namely calculated from the points for the symbol placement $(0, 0, 1, 0, n)$. Thus we are discussing only the symbol placements $(0, 0, 1, 0, 2)$ and $(0, 0, 1, 0, 3)$.

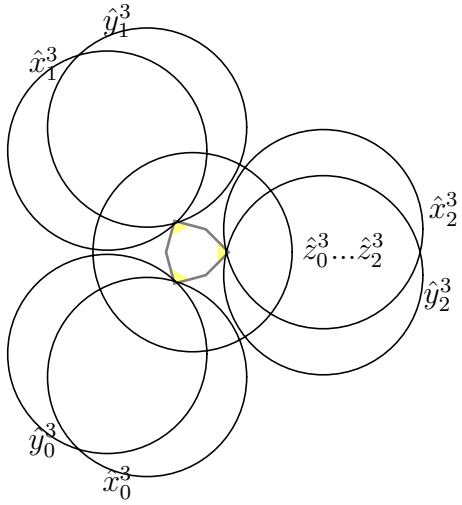
The following two pictures show the graph and the labeling for the symbol placement $(0, 0, 1, 0, 3)$.

*that means planar and embedded in the plane

Gadget graph **... labeling**



This picture shows the gadget graph and the symbol for the same symbol placement.



Note that this picture is drawn in a much smaller scale then the left picture. The small hexagon in the center is the same symbol than in the left picture.

The situation for the symbol placement $(0,0,0,0,2)$ is quite the same. The only difference is the positioning of the disks and the differences are too small to draw them correctly. The precise coordinates are given below.

for the symbol placement $(0,0,1,0,2)$	and for the symbol placement $(0,0,1,0,3)$
x_0=(.1090000,−.1887935)	
x_1=(.1090000, .1887935)	
x_2=(−.2180000, 0)	
y_0=(.2180000,−.3775871)	
y_1=(.2180000, .3775871)	
y_2=(−.4360000, 0)	
z_0=(−.5450000, .9439677)	
z_1=(−.5450000,−.9439677)	
z_2=(1.0900000, 0)	
$\overline{x}_0$=(−1.3680804,−3.7587700)	$\overline{x}_0$=(−1.3680805,−3.7587704)
$\overline{x}_1$=(−2.5711501, 3.0641774)	$\overline{x}_1$=(−2.5711504, 3.0641777)
$\overline{x}_2$=(3.9392315, .6945928)	$\overline{x}_2$=(3.9392309, −.6945927)
$\overline{y}_0$=(−2.5711501,−3.0641774)	$\overline{y}_0$=(−2.5711504,−3.0641777)
$\overline{y}_1$=(−1.3680804, 3.7587700)	$\overline{y}_1$=(−1.3680805, 3.7587704)
$\overline{y}_2$=(3.9392315, −.6945928)	$\overline{y}_2$=(3.9392309, −.6945927)
$\overline{z}_0$=(0,0)	
$\overline{z}_1$=(0,0)	
$\overline{z}_2$=(0,0)	

The points z_0, z_1, z_2 are exactly the same points as in the symbol for the symbol placement. The points $x_0, x_1, x_2, y_0, y_1, y_2$ lie somewhere inside the triangle $z_0 z_1 z_2$. The points $\overline{z}_0, \overline{z}_1, \overline{z}_2$ lie exactly at $(0, 0)$. The exact positioning of $\overline{x}_0, \overline{x}_1, \overline{x}_2, \overline{y}_0, \overline{y}_1, \overline{y}_2$ is a little bit tricky and needs further explanation.

First of all these points are very close to so called ideal points $x'_0, x'_1, x'_2, y'_0, y'_1, y'_2$. The distance to these points is always $\leqslant 0.00000052$ no matter whether we consider the symmetric of the symmetric gadget. Later this will be referred to as ***fact (a)***. Thus we would not notice any difference in the drawings. The ideal points are much easier to describe by using polar coordinates. The points all have a distance of 4 to the origin and the polar angles are $\frac{25}{18}\pi, \frac{13}{18}\pi, \frac{10}{18}\pi, \frac{23}{18}\pi, \frac{11}{18}\pi, -\frac{1}{18}\pi$. Thus the triangles x'_0, x'_1, x'_2 as well as y'_0, y'_1, y'_2 are regular. Next we need to define so called witness points $x''_0, x''_1, x''_2, y''_0, y''_1, y''_2$ which have the same polar angles but are exactly 0.99999999 far apart from the origin.

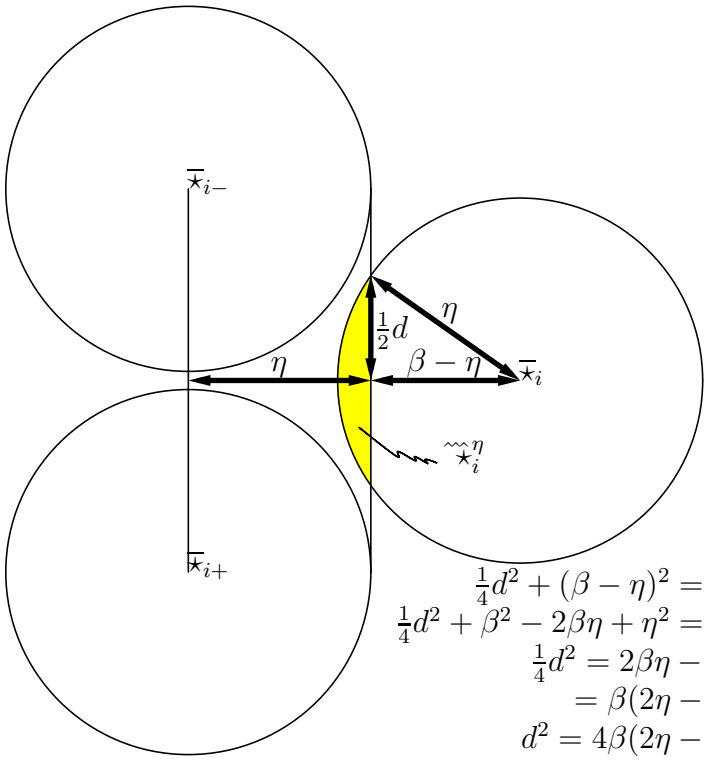
For convenience let $i+ := i+1 \bmod 3$ and $i- := i-1 \bmod 3$ for any $i \in \{0, 1, 2\}$. Furthermore let $\star$ represents either the letter x or y. For any η let $\hat{\star}_i^\eta$ be the closed disk centered at $\overline{\star}_i$ with radius η. Next let $\hat{\hat{\star}}_i^\eta$ denote the convex hull of $\hat{\star}_{i+}^\eta \cup \hat{\star}_{i-}^\eta$. Finally let $\hat{\hat{\hat{\star}}}_i^\eta := \hat{\star}_i^\eta \cap \hat{\hat{\star}}_i^\eta$. Then we have ***fact (b)*** stating that the diameter of $\hat{\hat{\hat{\star}}}_i^\eta$ is $\leqslant 0.0054$ for any $\eta \leqslant 3.0000001$. Then we claim that there are only two cases.

case (i) $\star''_i \in \mathrm{int}(\hat{\hat{\hat{\star}}}_i^3)$

case (ii) $\hat{\hat{\hat{\star}}}_i^\eta = \{\}$ holds even for $\eta = 3.0000001$.

Now we claim that for the symmetric gadget case (i) holds for any $\star_0, \star_1, \star_2$ and for the asymmetric gadget case (i) holds for any $\star_0, \star_1$ and case (ii) holds for any $\star_2$. You can verify these facts by doing some easy calculations on the numbers given above, but this would be a little bit annoying since there are so many numbers. Instead we wrote a program designed for the GNU bc interpreter which does the job.* In order to discuss the question how a disk $\hat{\star}_i^\eta$ intersects $\hat{\hat{\star}}_i^\eta$ we introduce a parameter β which is defined as the distance of $\overline{\star}_i$ to the line through $\overline{\star}_{i-}$ and $\overline{\star}_{i+}$. Assuming that the disks are located nearly in the way shown in the following picture (which will be the case since fact (a)) the diameter of the intersection (drawn yellow) will be $= d$.

*By the way as already pointed out in footnote 1 on page 26 we have also used a certain GNU bc program to calculate all the coordinates printed in this text.



$$\begin{aligned}
\tfrac{1}{4}d^2 + (\beta-\eta)^2 &= \eta^2 \\
\tfrac{1}{4}d^2 + \beta^2 - 2\beta\eta + \eta^2 &= \eta^2 \\
\tfrac{1}{4}d^2 &= 2\beta\eta - \beta^2 \\
&= \beta(2\eta - \beta) \\
d^2 &= 4\beta(2\eta - \beta)
\end{aligned}$$

In order to check fact (b) the program has to verify fact (a) which will be done at lines 89–91 and it has to verify that $4\beta(2\eta - \beta) \leqslant 0.0054^2$ holds which will be done at lines 43–45 and 60–62.

program

```
1  /* this program is for the GNU bc version 1.03 (versions 1.04–1.05 do not work)
2     invoked with "bc -l" because we need the
3     math library for trigonometric functions
4     a() is arc tan
5     s() is sine
6     c() is cosine. */
7
8  scale=20 /* scale determines the precision in terms
9              of number of digits after the decimal point */
10
11 define polar(v[],phi,radius) {
12  v[0]=radius*c(phi*a(1)/45)
13  v[1]=radius*s(phi*a(1)/45)
14 }
15
16 ipointradius=0.99999999
17 zz=polar(iy2[],-10,4)            /* y'_2 */
18 zz=polar(ix2[], 10,4)            /* x'_2 */
19 zz=polar(iy1[],110,4)            /* y'_1 */
20 zz=polar(ix1[],130,4)            /* x'_1 */
21 zz=polar(iy0[],230,4)            /* y'_0 */
22 zz=polar(ix0[],250,4)            /* x'_0 */
```

```
zz=polar(wy2[],-10,ipointradius) /* y''_2 */
zz=polar(wx2[], 10,ipointradius) /* x''_2 */
zz=polar(wy1[],110,ipointradius) /* y''_1 */
zz=polar(wx1[],130,ipointradius) /* x''_1 */
zz=polar(wy0[],230,ipointradius) /* y''_0 */
zz=polar(wx0[],250,ipointradius) /* x''_0 */

scale=90 /* ok we've done the trigonometry,
   from now we will be as exact as possible. */

beta=5.999999
diameter=0.0054
eta=3.0000001
distortion=0.00000052

define sqr(x) {
 return (x*x)
}

if (sqr(diameter)<4*beta*(2*eta-beta)) {
 "diameter alert" /* for fact (b) */
}

define sub(ba[],b[],a[]) {
 ba[0]=b[0]-a[0]
 ba[1]=b[1]-a[1]
 return (sqr(ba[0])+sqr(ba[1]))
}

define sidewise(a[],b[],c[]) {
 auto ba[],ca[],v,w
 zz=sub(ca[],c[],a[])
 w=sub(ba[],b[],a[]) /* w := ||b - a||^2 */
 v=sqr(ca[0]*ba[1]-ca[1]*ba[0])
 /* sqrt(v/w) is the distance of the point
      c to the line through a and b */
 if (v<w*sqr(beta)) {
  "diameter alert" /* for fact (b) */
 }
 if (v>w*sqr(2*eta)) {
  "safely not touching" /* for case (ii) */
 }
}

define inside(a[],b[],c[],i[]) { /* for case (i) */
 auto ic[],ia[],ba[],v,w
 /* the numeric constant 10^-15 here and
      elsewhere will mask rounding errors */
```

```
 if (sub(ic[],i[],c[])+10^-15>=9) {
  return (0) /* witness may not be inside circle */
 }
 zz=sub(ia[],i[],a[])
 w=sub(ba[],b[],a[]) /* w := ||b − a||^2 */
 v=sqr(ia[0]*ba[1]-ia[1]*ba[0])
 /* sqrt(v/w) is the distance of the point
    i to the line through a and b */
 if (v+10^-15<w*9) {
  "witness is safely inside intersection"
 }
}

define check(a[],b[],c[],i[],e[]) {
 auto z[]
 zz=sidewise(a[],b[],c[])
 zz=inside(a[],b[],c[],i[])
 if (sub(z[],c[],e[])+10^-15>sqr(distortion)) {
  " alert, to much distortion" /* for fact (a) */
 }
 print "\n"
}

define cartesic(v[],x,y) {
 v[0]=x
 v[1]=y
}

/* asymmetric case */
zz=cartesic(ax0[],-1.3680804,-3.7587700) /* x_0 */
zz=cartesic(ax1[],-2.5711501, 3.0641774) /* x_1 */
zz=cartesic(ax2[], 3.9392315,  .6945928) /* x_2 */
zz=cartesic(ay0[],-2.5711501,-3.0641774) /* y_0 */
zz=cartesic(ay1[],-1.3680804, 3.7587700) /* y_1 */
zz=cartesic(ay2[], 3.9392315, -.6945928) /* y_2 */
/* symmetric case */
zz=cartesic(sx0[],-1.3680805,-3.7587704) /* x_0 */
zz=cartesic(sx1[],-2.5711504, 3.0641777) /* x_1 */
zz=cartesic(sx2[], 3.9392309,  .6945927) /* x_2 */
zz=cartesic(sy0[],-2.5711504,-3.0641777) /* y_0 */
zz=cartesic(sy1[],-1.3680805, 3.7587704) /* y_1 */
zz=cartesic(sy2[], 3.9392309, -.6945927) /* y_2 */

print "asymmetric case\n";
" x0 "; zz=check(ax1[],ax2[],ax0[],wx0[],ix0[])
" x1 "; zz=check(ax2[],ax0[],ax1[],wx1[],ix1[])
" x2 "; zz=check(ax0[],ax1[],ax2[],wx2[],ix2[])
" y0 "; zz=check(ay1[],ay2[],ay0[],wy0[],iy0[])
```

```
" y1 "; zz=check(ay2[],ay0[],ay1[],wy1[],iy1[])
" y2 "; zz=check(ay0[],ay1[],ay2[],wy2[],iy2[])
print "symmetric case\n";
" x0 "; zz=check(sx1[],sx2[],sx0[],wx0[],ix0[])
" x1 "; zz=check(sx2[],sx0[],sx1[],wx1[],ix1[])
" x2 "; zz=check(sx0[],sx1[],sx2[],wx2[],ix2[])
" y0 "; zz=check(sy1[],sy2[],sy0[],wy0[],iy0[])
" y1 "; zz=check(sy2[],sy0[],sy1[],wy1[],iy1[])
" y2 "; zz=check(sy0[],sy1[],sy2[],wy2[],iy2[])

quit
```

output

```
asymmetric case
 x0 witness is safely inside intersection
 x1 witness is safely inside intersection
 x2 safely not touching
 y0 witness is safely inside intersection
 y1 witness is safely inside intersection
 y2 safely not touching
symmetric case
 x0 witness is safely inside intersection
 x1 witness is safely inside intersection
 x2 witness is safely inside intersection
 y0 witness is safely inside intersection
 y1 witness is safely inside intersection
 y2 witness is safely inside intersection
```

3.9 The reduction

Now we have all ingredients together needed to calculate an instance of the Fréchet metric problem from an instance of the selection problem.

First we put a sufficiently large triangle A around all symbols corresponding to the symbol placements in our ensemble. For each symbol placement in our ensemble we use the corresponding gadget described above which is a labeled triangulated plane graph inside of A. The outside of this graph forms a triangle. All these graphs together with the corners of A constitute a graph which is not fully triangulated and we triangulate it arbitrarily. Each corner of A we assign a disk of radius 3 centered at this vertex.

Altogether we get a labeled triangulated plane graph — let us call it G. Each of its vertices are points inside the triangle A, and the corners of A are vertices, too. Now we define $f\colon A \to \mathbb{R}^2$ such that f maps each vertex of G to the center of its assigned disk and such that f is affine on each triangle in G.

Now we define g to be the identity mapping on A and let $\varepsilon := 3$. Thus we have simplicial objects f, g and an $\varepsilon > 0$ and we are able to ask whether $\delta_F(f,g) \geqslant \varepsilon$ holds or not. This is our instance of the decision problem for the Fréchet metric.

Let us consider the decision problem again. Because g is the identity mapping on A it holds

$$\delta_F(f,g) = \inf_{\sigma: A \xrightarrow{\sim} A} \sup_{x \in A} ||f(x) - \sigma(x)||.$$

Now the question is whether $\forall \eta > 3\ \exists \sigma \colon A \xrightarrow{\sim} A\ \forall x \in A : ||f(x) - \sigma(x)|| \leqslant \eta$ holds.

In the following sections we will show that this is the case if and only if the instance of the selection problem is solvable. To be more precise we will show that

1. If the instance is solvable then $\exists \sigma \colon A \xrightarrow{\sim} A\ \forall x \in A : ||f(x) - \sigma(x)|| \leqslant \eta$ holds even for $\eta = 3$.

2. If it is not solvable then the same does not hold for $\eta = 3.0000001$.

We will show the first statement in the following section and the second statement in section 3.11 on page 48.

3.10 How the gadgets work

In this section we will show that if the instance of the selection problem is solvable then there exists a $\sigma \colon A \xrightarrow{\sim} A$ *with* $\forall x \in A : \delta_X\big(f(x), \sigma(x)\big) \leqslant 3$.

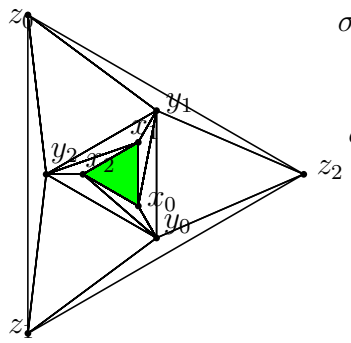
For a vertex x we would like to coin the term ***disk*** x for what is denoted by $\hat{x}^3$ in section 3.8 on page 33 which is the closed disk of radius 3 centered at $\overline{x}$. For an edge xy we would like to coin the term ***tube*** xy to be the convex hull of the disks x and y. These terms we will use in this sense only in this section. In the next section we will use the same terms in a slightly different manner.

Let therefore $\mathcal{A}$ be a selection solving the given instance of the selection problem. First we will show three different possible ways for such a σ to operate on a single *symmetric* gadget.

3.10.1 The three ways

way $\star_2$

The following pictures show the plane graph for a gadget again and the image of its edges. The dotted circles indicate the labeling of the graph.



plane gadget graph

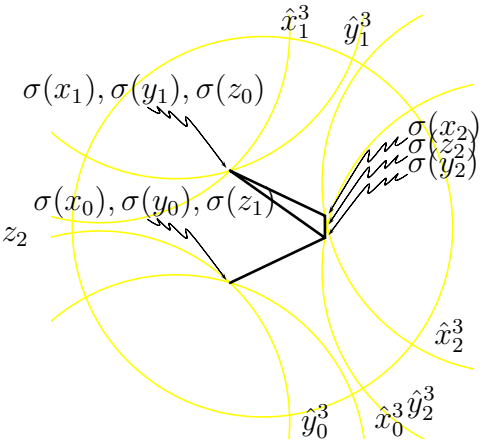


image of plane gadget graph under σ

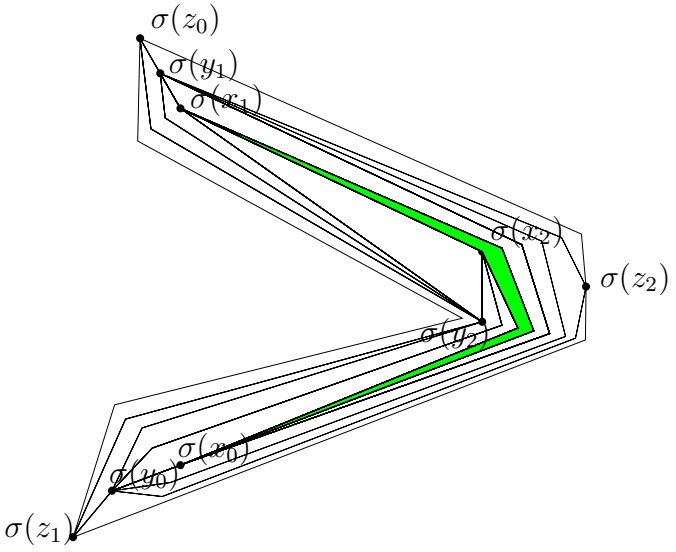
The invisible details of the construction are

$$\begin{aligned} \sigma(x_1), \sigma(y_1), \sigma(z_0) &\in \operatorname{int}(\hat{x}_1^3 \cap \hat{y}_1^3) \\ \sigma(x_0), \sigma(y_0), \sigma(z_1) &\in \operatorname{int}(\hat{x}_0^3 \cap \hat{y}_0^3) \\ \sigma(x_2) &\in \operatorname{int}(\hat{\hat{x}}_2^3) \\ \sigma(y_2) &\in \operatorname{int}(\hat{\hat{y}}_2^3) \end{aligned}$$

This is possible since case (i) on page 35 holds for x_2, y_2.

The image seems to consist only of 4 lines. This is, however, not the truth. You have to imagine each line as a bundle of curves which are so close together that it would not be possible to draw this true to scale. In fact even the whole green shaded triangle $x_0x_1x_2$ to the left is squeezed into these thick lines you see to the right.

In order to see that there is in fact an orientation preserving homeomorphism which squeeze the lines together in the way described before we would like to show a drawing of the image which is not true to scale but combinatorial correct.



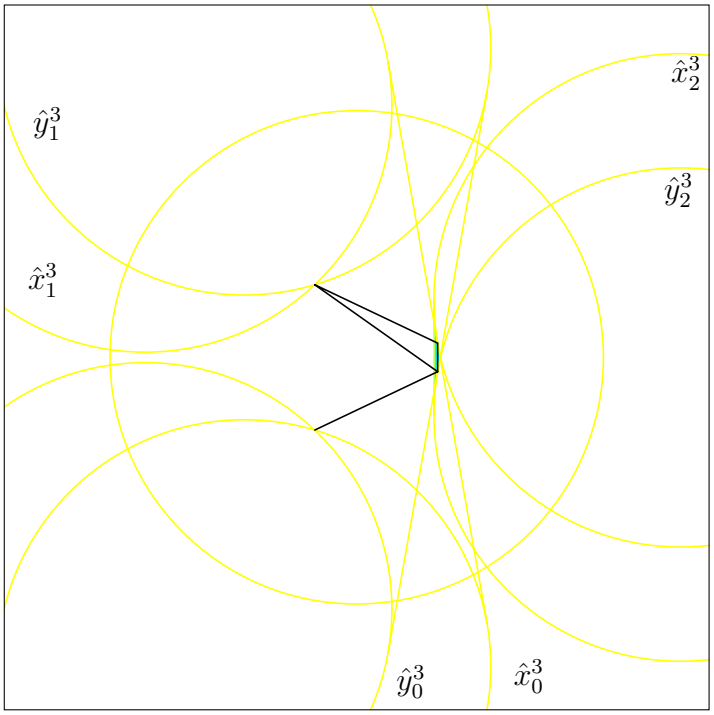
We say that an edge uv of the plane gadget graph is ***σ-monotone*** if and only if the image under σ is monotone in the direction of the line trough $\sigma(u)$ and $\sigma(v)$.

Our mapping σ will have the following properties. **(i)** It maps each σ-monotone edge uv in the following way. Let a be an affine mapping from the oriented line segment uv to the oriented line segment $\sigma(u)\sigma(v)$. For each point p on the line segment uv we consider the line going through $a(p)$ perpendicular to the line going through $\sigma(u)$ and $\sigma(v)$. Since uv is σ-monotone there will be exactly one intersection point between this line and $\sigma(uv)$ and then $\sigma(p)$ is that point. **(ii)** Each edge uv of the plane gadget graph is mapped onto a polygonal chain. **(iii)** For each image of a face of the plane gadget graph there is a triangulation using only the vertices of the simple polygon bounding the image such that σ is affine on each triangle of this triangulation.

Note that property (i) and (iii) do not contradict. We want to show that $||f(p) - \sigma(p)|| \leqslant 3$ holds for all points p inside of the triangle $z_0z_1z_2$. By property (ii) and (iii) we only have to show this for the corners of the polygonal chains constituting the images of the edges of the plane gadget graph.

By property (i) for each σ-monotone edge uv this is easy since we have only to check whether $\sigma(u)$ lies in the disk u and $\sigma(v)$ lies in the disk v and whether $\sigma(uv)$ lies inside the tube uv. Fortunately the σ we consider in what we have entitled "way $\star_2$" so far has the property that all edges are σ-monotone.

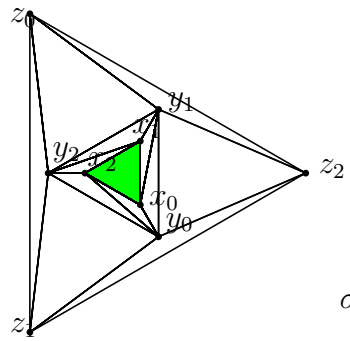
Now it is necessary to investigate the geometry of the mapping in more detail. The following picture shows the image again together with the disks and the tubes x_2y_2, y_0y_1 and x_0x_1. The cyan shaded triangle is the intersection of these tubes. Note that the line segment between and including $\sigma(x_2)$ and $\sigma(y_2)$ lies inside of the interior of this intersection. Note furthermore that the whole image of the plane gadget graph under σ is contained in the tubes y_0y_1 and x_0x_1.



It remains to show that for each edge uv of the plane gadget graph $\sigma(u)$ lies in the disk u and $\sigma(v)$ lies in the disk v and $\sigma(uv)$ lies inside the tube uv. You may verify that step by step for each edge and mark each edge with a pencil you have checked so far. See appendix on page 73.

way $\star_1$

Again we show the plane graph for a gadget and the image of its edges.



plane gadget graph

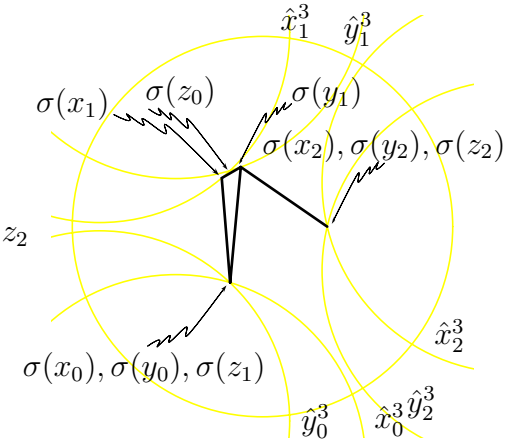


image of plane gadget graph under σ

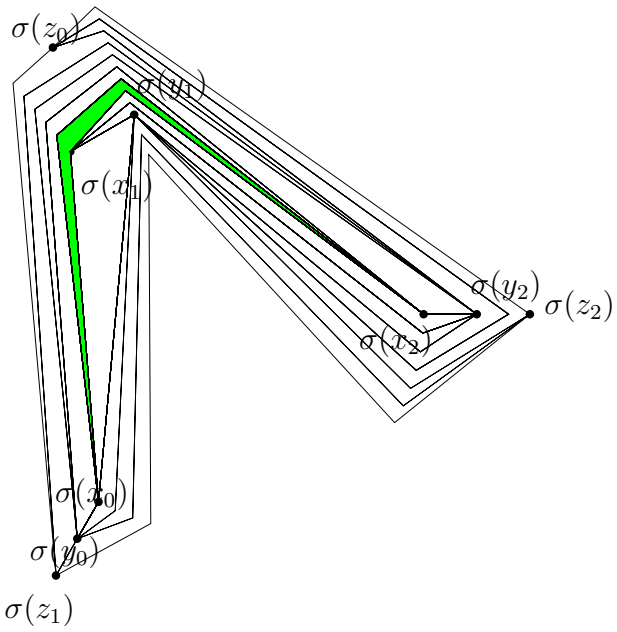
The invisible details of the construction are

$$\begin{aligned} \sigma(x_0), \sigma(y_0), \sigma(z_1) &\in \operatorname{int}(\hat{x}_0^3 \cap \hat{y}_0^3) \\ \sigma(x_2), \sigma(y_2), \sigma(z_2) &\in \operatorname{int}(\hat{x}_2^3 \cap \hat{y}_2^3) \\ \sigma(x_1) &\in \operatorname{int}(\hat{\hat{x}}_1^3) \\ \sigma(y_1) &\in \operatorname{int}(\hat{\hat{y}}_1^3) \end{aligned}$$

This is possible since case (i) on page 35 holds for x_1, y_1.

Note that the line segment between and including $\sigma(x_1)$ and $\sigma(y_1)$ lies inside of the interior of the intersection of the tubes x_1y_1, y_2y_0 and x_2x_0. Note furthermore that the whole image of the plane gadget graph under σ is contained in the tubes y_2y_0 and x_2x_0.

Again we show a drawing of the image which is not true to scale but combinatorial correct.



Again we want to show that $||f(p) - \sigma(p)|| \leqslant 3$ holds for all points p inside of the triangle $z_0z_1z_2$. Again we only have to show this for the corners of the polygonal chains constituting the images of the edges of the plane gadget graph.

Again for each σ-monotone edge uv this is easy. The σ we consider here in what we have entitled "way $\star_1$" has the property that all edges except y_1z_0 are σ-monotone.

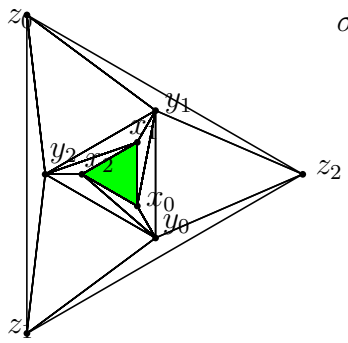
Since y_1z_0 is not monotone we have some freedom to choose σ and we assume that all points on the line segment y_1z_0 which are more far apart from z_0 than 10^{-42} are mapped * into the interior of $\hat{\hat{y}}_1^3$. Then it is easy to see that $||f(p)-\sigma(p)|| \leqslant 3$ holds for all points p on the line segment z_0y_1.

It remains to show that for each edge uv except z_0y_1 of the plane gadget graph $\sigma(u)$ lies in the disk u and $\sigma(v)$ lies in the disk v and $\sigma(uv)$ lies inside the tube uv. See appendix on page 75.

way $\star_0$

Again we show the plane graph for a gadget and the image of its edges.

*We have to introduce an artificial (redundant) corner on the polygonal chain constituting the image of y_1z_0 in order to keep the mapping affine on each segment.



plane gadget graph

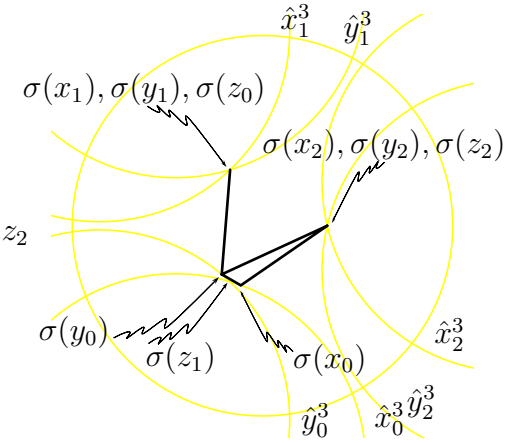


image of plane gadget graph under σ

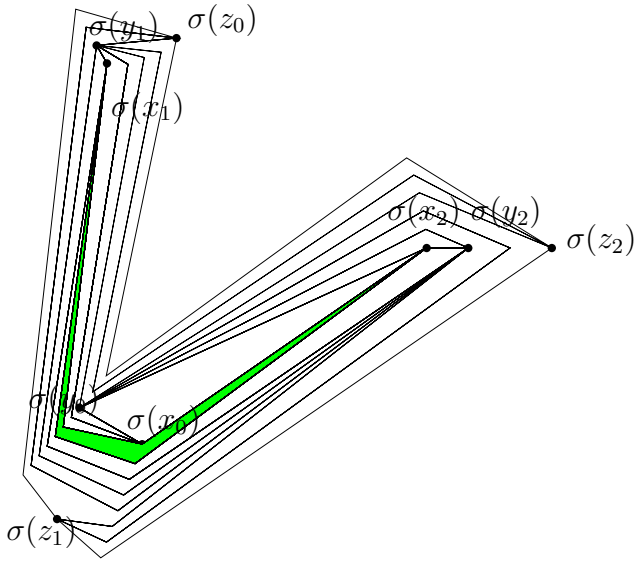
The invisible details of the construction are

$$\begin{aligned} \sigma(x_1), \sigma(y_1), \sigma(z_0) &\in \operatorname{int}(\hat{x}_1^3 \cap \hat{y}_1^3) \\ \sigma(x_2), \sigma(y_2), \sigma(z_2) &\in \operatorname{int}(\hat{x}_2^3 \cap \hat{y}_2^3) \\ \sigma(x_0) &\in \operatorname{int}(\hat{\hat{x}}_0^3) \\ \sigma(y_0) &\in \operatorname{int}(\hat{\hat{y}}_0^3) \end{aligned}$$

This is possible since case (i) on page 35 holds for x_0, y_0.

Note that the line segment between and including $\sigma(x_0)$ and $\sigma(y_0)$ lies inside of the interior of the intersection of the tubes x_0y_0, y_1y_2 and x_1x_2. Note furthermore that the whole image of the plane gadget graph under σ is contained in the tubes y_1y_2 and x_1x_2.

Again we show a drawing of the image which is not true to scale but combinatorial correct.



Again we want to show that $||f(p) - \sigma(p)|| \leqslant 3$ holds for all points p inside of the triangle $z_0z_1z_2$. Again we only have to show this for the corners of the polygonal chains constituting the images of the edges of the plane gadget graph.

Again for each σ-monotone edge uv this is easy. The σ we consider here in what we have entitled "way $\star_0$" has the property that all edges except y_2z_0, y_0z_1 and z_0z_1 are σ-monotone.

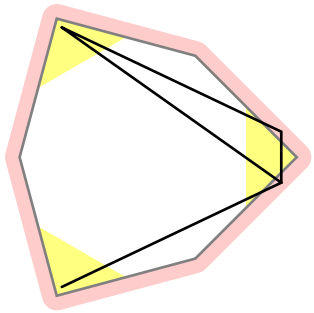
Since y_2z_0 and y_0z_1 are not monotone we have some freedom to choose σ and we assume that all points on the line segment y_2z_0 which are more far apart from z_0 than 10^{-42} are mapped into the interior of $\hat{\hat{y}}_2^3$. Analogously we assume that all points on the line segment y_0z_1 which are more far apart from z_1 than 10^{-42} are mapped into the interior of $\hat{\hat{y}}_0^3$.* Then it is easy to see that $||f(p) - \sigma(p)|| \leqslant 3$ holds for all points p on the line segments y_2z_0 and y_0z_1.

On the other hand, all points p of the line segment z_0z_1 are mapped into z_0^3 and thus $||f(p) - \sigma(p)|| \leqslant 3$ is trivial.

It remains to show that for each edge uv except y_2z_0, y_0z_1 and z_0z_1 of the plane gadget graph $\sigma(u)$ lies in the disk u and $\sigma(v)$ lies in the disk v and $\sigma(uv)$ lies inside the tube uv. See appendix on page 77.

3.10.2 Taking all together

The following picture shows the image of the plane gadget graph under σ as referred to as way $\star_2$ drawn together true to scale with the symbol.



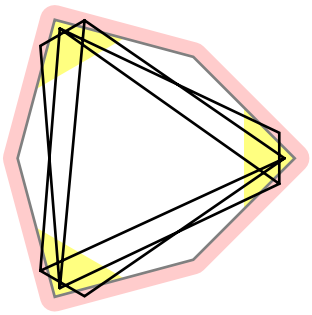
All three ways shown so far look like a big V with its base near to one of the yellow vertices of the symbol which will be subsequently referred to as the ***exposed vertex***. That is z_2 for way $\star_2$, z_0 for way $\star_1$ and z_1 for way $\star_0$. For the asymmetric gadget the situation is quite the same with the only exception that way $\star_2$ does not work (because then case (i) on page 35 does not hold for x_2, y_2). Thus for any symbol and for any of its yellow vertices it is possible to map the plane gadget graph of the symbol in a way shown above as a big V with the exposed vertex on this yellow vertex.

Now we are able to construct the whole σ for the given instance of the selection problem. Since the instance is solvable there exist a complete and feasible selection $\mathcal{A}$. For any symbol we map the plane gadget graph of the symbol such that this

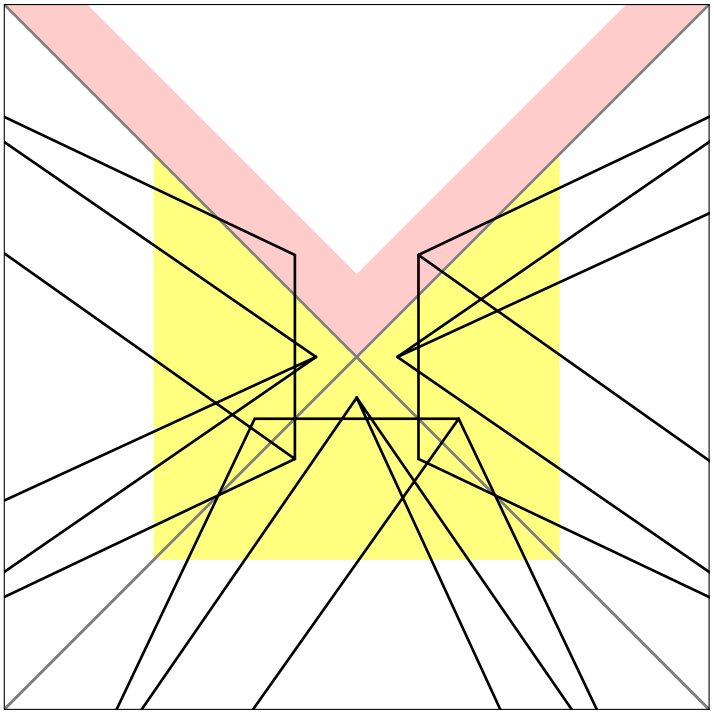
*See footnote 1 on page 44

yellow vertex of the symbol which is in $\mathcal{A}$ will be the exposed vertex. This is because $\mathcal{A}$ is complete.

Furthermore the images of the plane gadget graphs will not intersect. In order to see that we draw all three ways simultaneously (which will be shown in the following picture) and have a look on what happens if two gadgets touch each other.



In the following magnification of three touching gadgets you may see that the embeddings for different gadgets are not disturbing each other provided that exposed vertices are not touching. This can not be happen since $\mathcal{A}$ if feasible.



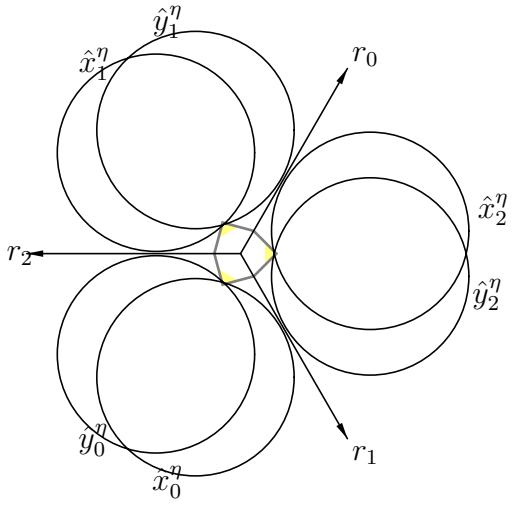
The mapping of the graph outside the gadgets will not cause further problems, since each of the outer vertices of the plane gadget graphs (which are vertices of the hexagons) will be moved only by a small amount and the remaining points outside will be moved only inside the interior of the $\hat{z}^3_{\dots}$.

3.11 How the gadgets do not work

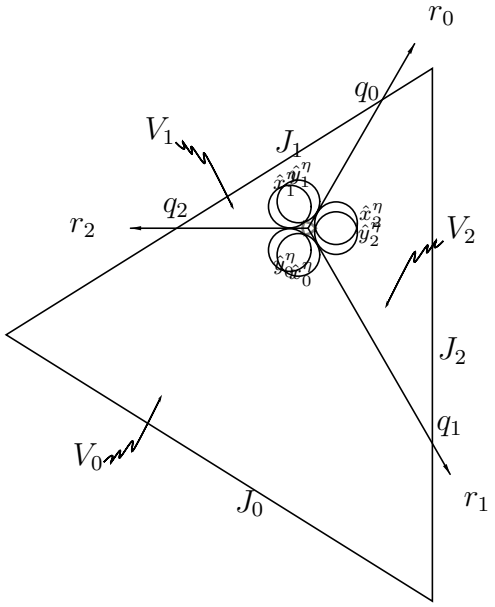
In this section we will show that if the instance of the selection problem is not solvable then there does not exist a $\sigma\colon A \xrightarrow{\sim} A$ *such that* $\forall x \in A : \delta_X\big(f(x), \sigma(x)\big) \leqslant 3.0000001$ *holds.*

Let $\eta := 3.0000001$.

Let us consider again a symmetric gadget more closely. For convenience let again $i_+ := i + 1 \bmod (3)$ and $i_- := i - 1 \bmod (3)$. Furthermore let $z := \overline{z}_0 = \overline{z}_1 = \overline{z}_2$ denote the ***center*** of the gadget. Now for each $i \in \{0, 1, 2\}$ let r_i be a ray from z through the midpoint between $\overline{x}_{i_-}$ and $\overline{y}_{i_+}$. The following figure should illustrate this.



Let q_i be that point where r_i leaves the triangle A. Let J denote the clockwise oriented boundary of A. Since r_0, r_1, r_2 are clockwise oriented the curve J can visit the points q_0, q_1, q_2 in this ordering. For each $i \in \{0, 1, 2\}$ let J_i be the (open) part on J from q_{i_+} to q_{i_-} (exclusively).



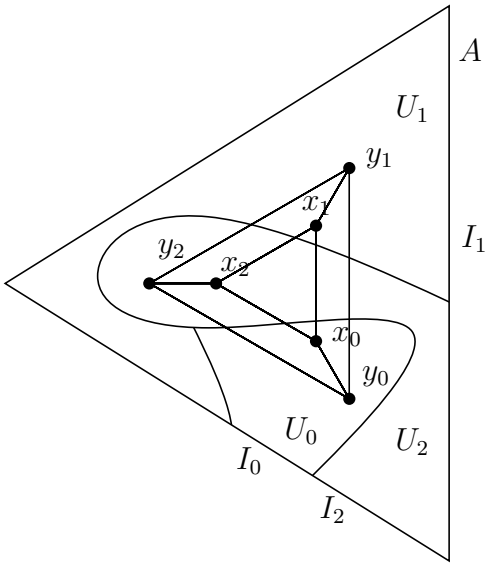
Without the rays r_0, r_1, r_2 the triangle A falls apart in three connected components V_0, V_1, V_2. Let V_i denote the component touching J_i for each i. The components V_0, V_1, V_2 are disjoint open subsets of A in the topology of the compact space A which are path-connected and bordered by Jordan curves.

The proof in the remainder of this section will be indirect. Let us therefore assume that there exists $\sigma\colon A \xrightarrow{\sim} A$ such that $\forall x \in A : \delta_X\big(f(x), \sigma(x)\big) \leqslant \eta$.

For a vertex x we would like to coin the term ***disk*** x to be be a closed disk of radius η centered at $\overline{x}$. For an edge xy we would line to coin the term ***tube*** xy to be the convex hull of the disks x and y. Finally the term ***curve*** xy should mean the image of the edge xy under σ. Obviously the curve xy lies inside of the tube xy.

Let U_i be the inverse image of V_i under σ. Let I_i the inverse image of J_i. Since $\sigma\colon A \xrightarrow{\sim} A$ is orientation preserving the parts I_0, I_1, I_2 will be encountered in this ordering which means in *clockwise* ordering.

The disks x_i and y_i and therefore also the tube x_iy_i and the curve x_iy_i lie inside of V_i. Their inverse image under σ is the edge x_iy_i and this edge has to lie in U_i therefore. Here is an example:



Let C_i denote the statement that every path from the line segment x_iy_i to I_i which lies wholly in U_i has to intersect the line segments $x_{k_-}x_{k_+}$ and $y_{k_-}y_{k_+}$. As an example in the drawing above C_0, C_1 are false whereas C_2 is true. Furthermore let P_i be the statement that every polygonal path does the same.

Lemma. Given an in A open set $O \subseteq A$ and two points $s, t \in O$. If there exists a path from s to t in O then there exists a polygonal path, too.

Proof. Let $\gamma\colon [0,1] \to O$ be a path from s to t. For each point $x \in [0,1]$ it holds $\gamma(x) \in O$. Since O is open, there exists an open disk D such that its intersection with A lies wholly in O. Since D is open and γ is continuous and $\gamma(x) \in D$ holds, it will be an open interval $U(x)$ around x such that $\gamma[U(x)] \subseteq D$ holds. Together with D and A also $D \cap A$ is convex. Now $\gamma[U(x)] \in D \cap A \subseteq O$ holds and therefore the convex hull of $\gamma[U(x)]$ lies in O. Since $[0,1]$ is compact there

exist finitely many intervals from $\{U(x) : x \in [0,1]\}$ which cover $[0,1]$. Let those intervals be sorted by starting points in increasing order denoted with $U_0, ..., U_n$. It will be $0 \in U_0$ and it will exists a $m \in \{0, ..., n\}$ with $1 \in U_m$. For each $i \in \{0, ..., m-1\}$ the intervals U_i and U_{i+1} will intersect. Let us choose now a $x_{i+1} \in U_i \cap U_{i+1}$. If we define finally $x_0 := 0$ and $x_{m+1} := 1$ then $x_i, x_{i+1} \in U_i$ will hold for all $i \in \{0, ..., m\}$. Since the convex hull of $\gamma[U_i]$ lies in O the line segment $\gamma(x_i)\gamma(x_{i+1})$ will lie in O. The polygonal chain $\gamma(x_0), ..., \gamma(x_{m+1})$ will do what we want. □

Claim. $C_0 \vee C_1 \vee C_2$

Proof.

For all $i \in \{0,1,2\}$ we have first that U_i is a path-connected open set in which x_i and I_i lies. By the lemma above there exists a polygonal curve α_i from x_i to a point x_i' on I_i. Let us fix ***this*** $\boldsymbol{\alpha_i}$.

In the following we will deal mainly with polygonal curves only, which may intersect or not. We want to characterize the possible intersections of two different polygonal chains. Let us define an ***intersection*** to be a pair of common sub curves of the corresponding polygonal curves which can not be enlarged. A ***crossing*** will be an intersection such that the curves exchange their sides. This notion makes only sense for closed curves or in the case that they do not begin or end within the intersection. Between two closed polygonal curves there can be only a finite number of intersections and among them only an even number of crossings.*

Let us consider now the set E of all vertices of all $\alpha_0, \alpha_1, \alpha_2$. We select now a point $\boldsymbol{p}$ in the interior of the triangle $x_0x_1x_2$, which does not lie on one or more of the finitely many lines running through two points of the finite set $E \cup \{x_0, x_1, x_2\}$. Thus p will not be contained in the curves $\alpha_0, \alpha_1, \alpha_2$ and the line segments px_0, px_1, px_2 will contain none of the vertices of those curves. In particular these line segments will intersect those curves only in finitely many points and these points will be no vertices on the curves. Thus any intersection between these line segments and those curves will be a crossing.

For each i let $\boldsymbol{\beta_i}$ now denote the curve which first runs straight from p to x_i and then continues with α_i. The part until but included x_i we want to call the ***start*** and the remaining part the ***ending***. The starts intersect each other only in p and the endings are nothing else than $\alpha_0, \alpha_1, \alpha_2$ and therefore disjoint.

Consider now the intersections of $\beta_0, \beta_1, \beta_2$ with each other. Except from p all intersections will be intersections of an ending of a β_i with a start of β_{i_-} or β_{i_+} and that means that α_i crosses the polygonal chain $x_{i_-}px_{i_+}$. Let σ_i be the number of this crossings.

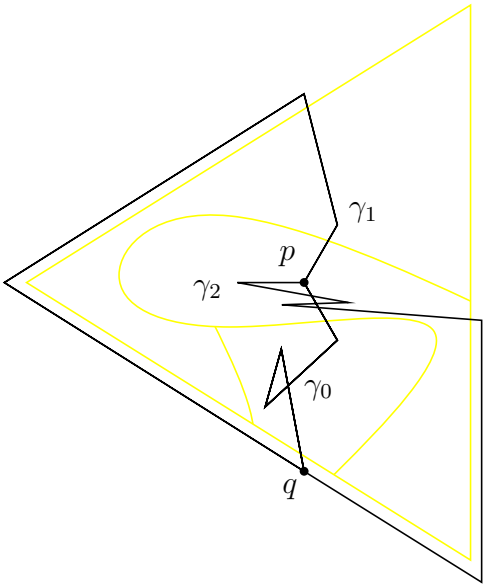
Thus we have $\sigma := \sigma_0 + \sigma_1 + \sigma_2$ crossings between $\beta_0, \beta_1, \beta_2$.

*If we would restrict this to simple curves this would follow directly from jordaens theorem. In fact we do not assume the curves to be simple but the statement still remains true even in that case. It may be proven by Jordan theorem again but the proof is more tricky in that case. On the other hand the statement is weaker since it deals only with polygonal curves and Jordan theorem is simpler to prove in that case.

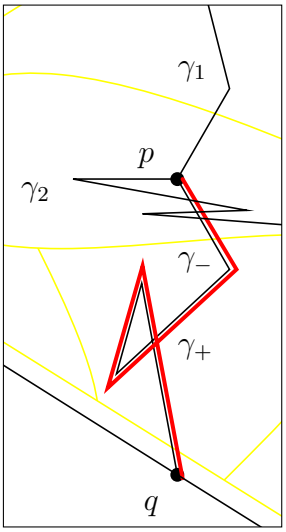
Notion For any curve c let c^{rev} denote the same curve backwards.
Claim. σ is odd.

Proof.

First we enlarge $\beta_0, \beta_1, \beta_2$ by adding curves outside of A meeting in a common endpoint q and intersecting only at q. Since their origins (i.e. the endpoints of $\beta_0, \beta_1, \beta_2$) are clockwise ordered they will meet in counter clockwise order. Let $\gamma_0, \gamma_1, \gamma_2$ denote the enlarged curves. They all starts from p in counter clockwise order they all meet in q in counter clockwise order. Except from p and q they all intersect only pairwise and σ times at all. Here is a picture:



Now you may already see the claim. For a rigid proof you should split γ_0 into two sides namely a left* one γ_+ and a right one γ_-. In the following magnification the left side γ_+ is shown highlighted.



*That means left when moving along γ_0 forward. Since in the picture above γ_0 goes mainly from upside down, this is a little bit confusing.

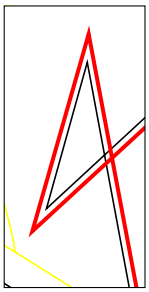
Since the curves are polygonal we may view γ_- and γ_+ as two curves which are very close to γ close enough that the number of intersections between them and the other curves are equal for both.

Let α denote the curve composed of γ_+ and γ_1^{rev}. Let β the one composed of γ_- and γ_2^{rev}. Now α and β will be closed polygonal curves. Thus they have to cross each other an even number of times. We can divide the crossings between α and β into

$2c_{00}$	crossings between γ_+ and γ_-
c_{02}	crossings between γ_+ and γ_2^{rev}
c_{10}	crossings between γ_1^{rev} and γ_-
c_{12}	crossings between γ_1^{rev} and γ_2^{rev}
0	crossings at p
1	crossings at q

where c_{ij} denotes the number of crossings between β_i and β_j for all $i \neq j$ and c_{00} denotes the number of self intersections of γ_0.

The following magnification shows how a self intersection of γ_0 generates exactly two crossings between γ_+ and γ_-.



Since $c_{02} + c_{10} + c_{12}$ expresses the total number of crossings between $\beta_0, \beta_1, \beta_2$ this is $= \sigma$. Therefore the total number of intersections between α and β calculates to be $2c_{00}+\sigma+1$. Since this number has to be even anyway and $2c_{00}+1$ is obviously odd σ has to be odd, too.

□

Since $\sigma_0 + \sigma_1 + \sigma_2 = \sigma$ is odd there must be a k for which σ_k is odd. Let us fix ***this k***.

Consider now the following curve composed of α_{k_+}, the piece of J from x'_{k_+} to x'_{k_-} and $\alpha_{k_-}^{\mathsf{rev}}$. This curve lies completely outside of U_k. By adding the polygonal chain $\gamma := x_{k_-} p x_{k_+}$ to the curve we get a closed curve. Let us call this closed curve Γ. Remember that every polygonal chain lying inside U_k can cross Γ only at γ.

Let α be now an arbitrary polygonal chain within U_k from a point x on the line segment $x_k y_k$ to a point x' on I_k. Let β now be the closed polygonal chain which starts at x_k with α_k then goes inside I_k anyhow from x'_k to x' then with α^{rev} to x then straight to x_k. The closed curves Γ and β can cross only an even number of times and all crossings between them are those between γ and α_k or α.

Since the number of crossings between α_k and γ is exactly σ_k and therefore odd the number of crossings between α and γ has to be odd, too.

Consider now the triangle $x_{k_-}px_{k_+}$. The points x_{k_-} and x_{k_+} lie inside U_{k_-} and U_{k_+}, respectively and thus not in U_k. Therefore α can not meet them. Furthermore the starting and ending points of α do not lie inside of this triangle. Together with the number of crossings between α and γ the number of the ones between α and the line segment $x_{k_-}x_{k_+}$ has to be odd, too.

Consider now the quadrangle $x_{k_-}x_{k_+}y_{k_+}y_{k_-}$. The line segments $x_{k_+}y_{k_+}$ and $y_{k_-}x_{k_-}$ lie inside U_{k_+} and U_{k_-}, respectively and therefore not in U_k. Therefore α can not intersect them. Furthermore the starting and ending points of α do not lie inside the quadrangle. Together with the number of crossings between α and the line segment $x_{k_-}x_{k_+}$ the number of the ones between α and the line segment $y_{k_-}y_{k_+}$ has to be odd, too.

The number of crossings between α and each of the line segments $x_{k_-}x_{k_+}$ and $y_{k_-}y_{k_+}$ is odd and therefore not zero in both cases. Therefore α has to intersect both line segments.

We have only supposed that α was an arbitrary polygonal chain from a point x on the line segment x_ky_k to a point x' on the curve I_k. We have observed that α has to intersect both line segments. This proves P_k.

Now we assume that C_k would not hold. Then there would exist a path from a point s on the line segment x_ky_k to a point t on I_k, which lies entirely in U_k and which does not intersect both line segments $x_{k_-}x_{k_+}$ and $y_{k_-}y_{k_+}$. Let L be one of these line segments which does not intersect the path. By the lemma on page 49 above there would exist a polygonal path from s to t which lies entirely in $U_k \setminus L$. This would contradict P_k. □

Claim. If C_k holds then the curve x_ky_k lies in each of the tubes x_ky_k, $x_{k_-}x_{k_+}$ and $y_{k_-}y_{k_+}$.

Proof. The curve lies in the tube x_ky_k anyway. For the rest of the proof let p be an arbitrary point on the curve x_ky_k. By $p \in V_k$ it holds $p \neq z$. Let therefore r be the ray starting at p moving straight away from z on the line through z and p. Let q be the point at which r leaves the triangle A.

The line segment pq will be entirely inside V_k. Its inverse image under σ will be a path from $\sigma^{-1}(p)$ to $\sigma^{-1}(q)$ inside of U_i. Now $\sigma^{-1}(p)$ lies on the line segment x_ky_k and $\sigma^{-1}(q)$ lies on I_k. Thus the path has to intersect $x_{k_-}x_{k_+}$ and $y_{k_-}y_{k_+}$. But that means that the line segment pq intersects the curves $x_{k_-}x_{k_+}$ and $y_{k_-}y_{k_+}$. The line segment pq is a part of r and the curves has to lie in the tubes $x_{k_-}x_{k_+}$ and $y_{k_-}y_{k_+}$. Therefore r has to intersect these tubes.

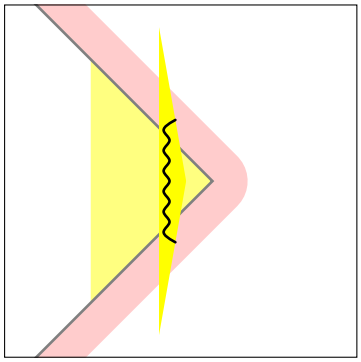
Let T be one of these tubes. Since r intersects T there must be a point $p' \in T$ on r. By construction of r the point p lies between z and every point on r. Thus p lies between z and p'. By $z, p' \in T$ follows $p \in T$ since T is convex.

□

Now we want to say a vertex z_k to be ***exposed*** if and only if C_k holds. Let $\mathcal{A}$ be the set of all exposed vertices of symbols in our ensemble which was the given

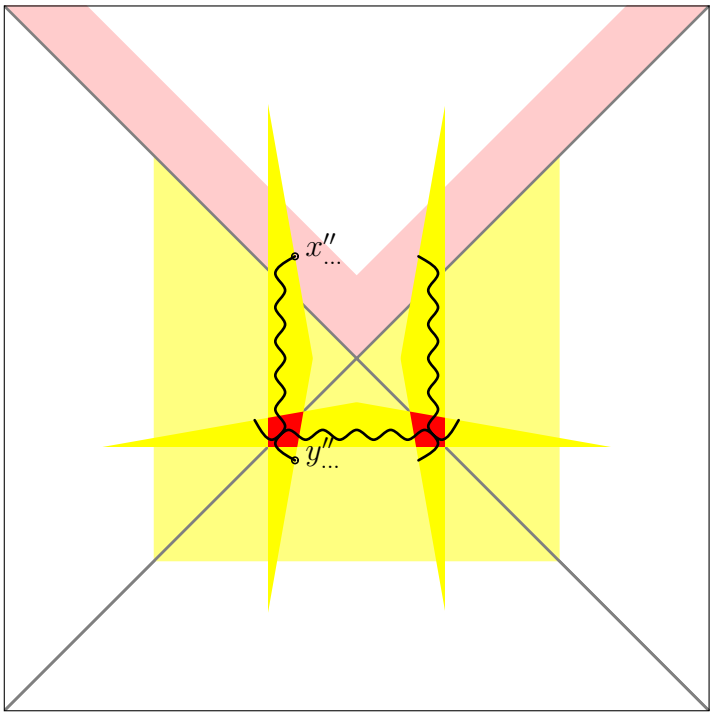
instance of the selection problem.

What we have learned so far in this section is that for $\eta = 3.0000001$ there must be an exposed vertex z_k —let us call this fact $(\star)$— and that for such a k the curve $x_k y_k$ has to lie inside of the tubes $x_k y_k$, $x_{k_-} x_{k_+}$ and $y_{k_-} y_{k_+}$. The intersection of these tubes is a small triangle which is shown true to scale in the following magnification of the symbol of a gadget for $k = 2$.



The wiggled line indicates how the curve $x_k y_k$ could look like. Its starting and ending points has to lie fairly exact at the positions drawn in the picture, anyway. This is because they have to lie in the intersection of the triangle with the disks x_k and y_k, respectively and that these intersections are very small. Additionally we can deduce that only the yellow vertices of a symbol can be exposed. Thus $\mathcal{A}$ is a selection. By $(\star)$ it is complete.

The following picture shows three touching gadgets.



There is some inaccuracy in these drawings. First note that the edges of the touching gadgets do not necessary match exactly instead they are only μ-close

where $\mu = 0.0000009$ as defined in page 27. So there is a positioning inaccuracy. But the inaccuracy is clearly smaller than what is visible on the drawing above. Second the beginnings and the endings of the wiggled lines are located in the drawing at some points $x''_{\ldots}$ and $y''_{\ldots}$ of the appropriate gadgets whereas the only thing which is sure is that these points are located in some intersections of a tube and a circle. These intersections have a diameter of $\leqslant 0.0054$ by fact (b) on page 35. In order to see that this will not cause any problems in the left gadget circles of radius 0.0055 are drawn around the points $x''_{\ldots}$ and $y''_{\ldots}$.

Now we can see that it is indeed impossible that for two gadgets touching at a common edge their exposed vertices are simultaneously at the same place, since otherwise two of the curves drawn as wiggled lines has to intersect in the red shaded quadrangle which forms the intersection of two triangles. Consider for example the left gadget and the lower one. The curve in the left one has to pass the quadrangle upside down and the curve for the lower one has to pass the quadrangle from left to right.

This makes $\mathcal{A}$ feasible. Thus our instance would be solvable. □

3.12 Remarks

As we have seen, the decision problem for the Fréchet metric is NP-hard. Next we could hope to find a polynomial time algorithm for an δ-approximate version of the decision problem. That is, we allow the algorithm to answer "I don't know" if the value of the distance is within $[\varepsilon - \varepsilon\delta, \varepsilon + \varepsilon\delta]$. But what we actually have proven is that for $\delta := 0.000000016$ the δ-approximate decision problem is NP-hard, too. To see this, choose $\varepsilon := 3.00000005$. Then $\varepsilon\delta < 0.000000049$ holds, hence $3 < \varepsilon - \varepsilon\delta$ and $\varepsilon + \varepsilon\delta < 3.0000001$. By our construction either $\delta_F(f, g) \leqslant 3$ or $\delta_F(f, g) \geqslant 3.0000001$ is obtained thus even a δ-approximate algorithm will always answer correctly to the instance created by our reduction.

3.13 Open problems

1. It is not known, whether the decision problem is in NP. It is not even known whether the decision problem is decidable even in an approximate sense. In fact, we have tried to prove the problem to be decidable for quite a long time, it seems to be an intriguing problem.

   Note that the problem to decide whether two four dimensional manifolds are homeomorphic is undecidable (due to [Mar58], but see [Hak] for a more readable introduction). Thus if we define δ_F to be $+\infty$ for objects with non homeomorphic domain we can say with good reasons that the decision problem for the Fréchet metric is undecidable for dimension $\geqslant 4$. Nobody knows what is between.

2. Furthermore one can hope that polynomial algorithms exist if we restrict the input to describe "simpler" figures.

(a) For boundaries of convex bodies in a natural parameterization this is easy because here the Fréchet distance equals to the Hausdorff distance of the bodies.

(b) Simple, i.e. non intersecting surfaces, that mean that the functions must be injective. For $X = I\!R^4$ this is NP-hard, too.* For $X = I\!R^3$ this is also probably NP-hard, but nothing is known about that.

(c) Surfaces in which the corners are far from each other. Nothing is known about that.

3. And finally one can hope to find a polynomial δ-approximate algorithm for some fixed δ. But after the remark mentioned above it is clear that the δ can not be arbitrarily small.

*Consider the mapping $(x, y) \mapsto (cx, cy, f_1(x), f_2(x))$ for a very small c, where $(x, y) \mapsto (f_1(x), f_2(x))$ describes the original (non simple) surface f constructed in our proof.