Chapter 4

Matching polygonal curves with respect to the Fréchet distance

In this chapter we will develop exact and approximation algorithms for the following problem:

Problem 4.1 (Matching problem – optimization version).

Given *two polygonal curves* $P, Q \in \mathcal{K}^0$.

Find *a translation* τ *such that* $\delta_F(\tau(P), Q)$ $(\tilde{\delta}_F(\tau(P), Q))$ *is as small as possible.*

To be more precise, we will provide an approximation algorithm in section 4.3 which does not necessarily compute the optimal transformation, but one that yields a Fréchet distance which differs from the optimum value by a factor of $(1 + \epsilon)$. To this end, we observe that it is easy to generalize the notion of a *reference point* to the Fréchet metric and apply the machinery of reference point based matching. The algorithm will run in time $\mathcal{O}(\epsilon^{-2}mn)$, where m and n denote the number of vertices of P and Q, respectively (c.f., Theorem 4.29 on page 40); it constitutes the first approximation algorithm for Problem 4.1. We conclude the section with a proof of the fact that there are no reference points for affine maps (c.f., Theorem 4.31 on page 41)

In section 4.2 we will consider the decision problem version of the exact matching problem:

Problem 4.2 (Matching problem – decision version).

Given *two polygonal curves* $P, Q \in \mathcal{K}^0$*, and* $\delta \geq 0$.

Decide, *whether there exists a translation* τ *such that* $\delta_F(\tau(P), Q) \leq \delta$ $(\tilde{\delta}_F(\tau(P), Q) \leq \delta)$.

We describe the first algorithm that solves the decision problem; it runs in $\mathcal{O}\big((mn)^3(m+n)^2\big)$ $(\mathcal{O}\big((mn)^3\big))$ time (c.f., Theorem 4.23 on page 37 and Theorem 4.24 on page 38). Efrat et al. [33] have independently developed an algorithm for the decision problem. However we should point out, that the runtime they achieve is by a factor mn slower than ours; furthermore their result is rather complicated and relies on complex data structures. Their work is based on [49], where an algorithm is presented that solves the decision problem for translations in a fixed direction in $\mathcal{O}((mn)^2(m+n))$ time.

4.1 Computing the Fréchet distance

Let us first describe how to *compute* the (weak) Fréchet distance between two polygonal curves, according to Alt and Godau, [13]. This will provide us with some of the main ingredients of our algorithm. For the remainder of this Part, $P : [0,m] \to \mathbb{R}^2$ and $Q : [0,n] \to \mathbb{R}^2$ will be polygonal curves, $\delta \geq 0$ is a fixed real parameter, and $\mathcal{T}_2$ denotes the set of planar translations. By slightly generalizing our notation, we will assume that a polygonal curve P is parametrized over $[0,n]$, where n is the number of segments of P, and $P|_{[i,i+1]}$ is affine for all $0 \leq i < n$.

A translation $\tau = \langle (x,y) \mapsto (x+\delta_x, y+\delta_y)\rangle \in \mathcal{T}_2$ can be specified by the pair $(\delta_x, \delta_y) \in \mathbb{R}^2$ of its *parameters*. The set of parameters of all translations in $\mathcal{T}_2$ is called the parameter space of $\mathcal{T}_2$, or *translation space* for short, and we identify $\mathcal{T}_2$ with its parameter space $\mathbb{R}^2$.

In the sequel we will use the notion of a *free space* which was introduced in [13]:

Definition 4.3 (Free space, Alt/Godau, [13]). *The set* $F_\delta(P,Q) := \{(s,t) \in [0,m] \times [0,n] \mid ||P(s) - Q(t)|| \leq \delta\}$*, or* F_δ *for short, denotes the* free space *of* P *and* Q*.*

Sometimes we refer to $[0,m] \times [0,n]$ as the *free space diagram*; the *feasible* points $p \in F_\delta$ will be called 'white' and the *infeasible* points $p \in [0,m] \times [0,n] - F_\delta$ will be called 'black' (for obvious reasons, c.f. Figure 4.1). Consider $[0,m] \times [0,n]$ as composed of the mn cells $C_{i,j} := [i-1,i] \times [j-1,j]$ $1 \leq i \leq n,\ 1 \leq j \leq m$. Then $F_\delta(P,Q)$ is composed of the mn free spaces for each pair of edges $F_\delta(P_{i-1}, Q_{j-1}) = F_\delta(P,Q) \cap C_{i,j}$. By L and R, we will denote the lower left and the upper right corner of F_δ, respectively, i.e., $L := (0,0)$, and $R := (m,n)$.

The following results from [13] describe the structure of the free space and link it to the problem of computing δ_F and $\tilde{\delta}_F$.

Lemma 4.4 (Properties of the free space, Alt/Godau, [13]).

1. The free space of two line segments is the intersection of the unit square with an affine image of the unit disk, i.e., with an ellipse, possibly degenerated to the space between two parallel lines.

2. For polygonal curves P and Q we have $\tilde{\delta}_F(P,Q) \leq \delta$, exactly if there exists a path within $F_\delta(P,Q)$ from L to R.

3. For polygonal curves P and Q we have $\delta_F(P,Q) \leq \delta$, exactly if there exists a path within $F_\delta(P,Q)$ from L to R which is monotone in both coordinates; such a path will be called *bi-monotone*.

Definition 4.5 (δ_F-path, $\tilde{\delta}_F$-path). *A path* π *in* F_δ *from* L *to* R *will be called a* $\tilde{\delta}_F$-path*. If* π *is also bi-monotone, it will be called a* δ_F-path.

For a proof of the Lemma we refer to [13]. Figure 4.1 shows polygonal curves P, Q, a distance δ, and the corresponding diagram of cells $C_{i,j}$ with the free space F_δ. Observe that the curve π as a continuous mapping from $\mathbb{S}^0$ to $[0,m] \times [0,n]$ directly gives feasible reparametrizations, i.e., two reparametrizations α and β, such that $\max_{t\in\mathbb{S}^0}||P(\alpha(t)) - Q(\beta(t))|| \le \delta$.

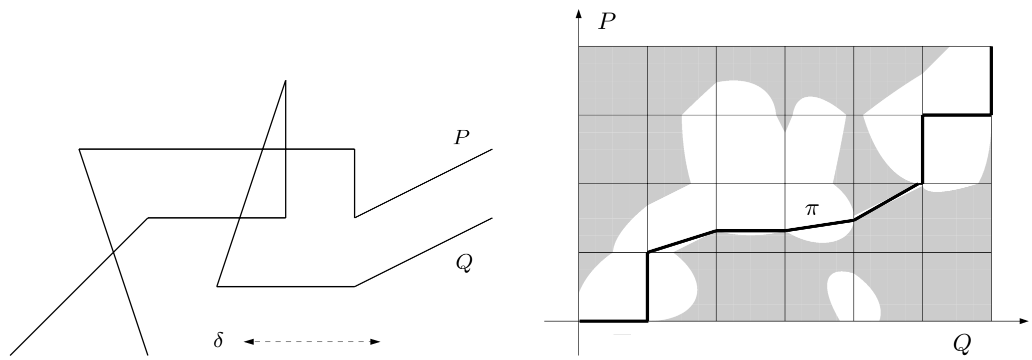

Figure 4.1: Two polygonal curves P and Q and their free space diagram for a given δ. An example δ_F-path π in the free space is drawn bold.

For $(i,j) \in \{1,\dots,m\} \times \{1,\dots,n\}$ let $L_{i,j} := \{i-1\} \times [a_{i,j}, b_{i,j}]$ ($B_{i,j} := [c_{i,j}, d_{i,j}] \times \{j-1\}$) be the left (bottom) line segment bounding $C_{i,j} \cap F_\delta$ (see Figure 4.2). The segment $\alpha_{i,j} := \{(i-1,y) \mid j-1 \le y < a_{i,j}\} \subseteq \bar{F}_\delta$ is called a *bottom-spike*, and the segment $\beta_{i,j} := \{(i-1,y) \mid b_{i,j} \le y < j\} \subseteq \bar{F}_\delta$ is called a *top-spike*. Likewise, the segment $\gamma_{i,j} := \{(x,j-1) \mid i-1 \le x < c_{i,j}\} \subseteq \bar{F}_\delta$ is called a *left-spike*, and the segment $\delta_{i,j} := \{(x,j-1) \mid d_{i,j} \le x < i\} \subseteq \bar{F}_\delta$ is called a *right-spike*. A bottom-spike $\alpha_{i,j}$ and a top-spike $\beta_{k,j}$ will be called *aligned*, if $a_{i,j} = b_{i,j}$. Likewise, a left-spike $\gamma_{i,j}$ and a right-spike $\delta_{k,j}$ will be called *aligned*, if $c_{i,j} = d_{i,j}$.

By induction it can easily be seen that those parts of the segments $L_{i,j}$ and $B_{i,j}$ which are reachable from L by a bi-monotone path in F_δ are also line segments. Using a dynamic programming approach one can compute them, and thus decide if $\delta_F(P,Q) \le \delta$. For details we refer to the proof of the following Theorem in [13]:

Theorem 4.6 (Computing the Fréchet distance (decision problem), Alt/Godau, [13]). One can decide in $\mathcal{O}(mn)$ time, whether $\delta_F(P,Q) \le \delta$ ($\tilde{\delta}_F(P,Q) \le \delta$).

As we have already mentioned, each (possibly clipped) ellipse in F_δ is the affine image of a unit disk. Each such 'ellipse' in F_δ varies continuously in δ; to be more precise, the cell boundaries $a_{i,j}(\delta)$, $b_{i,j}(\delta)$, $c_{i,j}(\delta)$, and $d_{i,j}(\delta)$ are continuous functions of δ. This implies that when δ is as small as possible, i.e., $\delta = \delta_F(P,Q)$ ($\delta = \tilde{\delta}_F(P,Q)$), all δ_F-paths ($\tilde{\delta}_F$-paths) in the diagram have to contain some (in fact at least two) of the extremal points of the free space, i.e., the extreme points of the cell boundaries: $a_{i,j}$, $b_{i,j}$, $c_{i,j}$, and $d_{i,j}$.

Definition 4.7 (Clamped path). *Let π be a $\tilde{\delta}_F$-path in F_δ. We call π* clamped in the *j*-th row (column) *if $a_{i,j} = b_{k,j}$ ($c_{j,i} = d_{j,k}$) for some $i \le k$, i.e., if $\alpha_{i,j}$ and $\beta_{k,j}$ ($\gamma_{j,i}$ and*

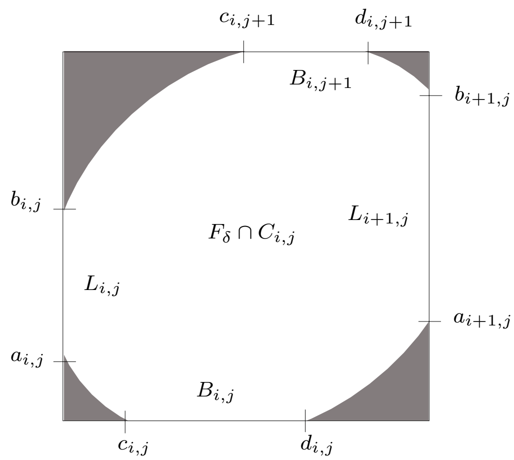

Figure 4.2: Intervals of the free space on the boundary of a cell.

$\delta_{j,k}$) are aligned, and $(i, a_{i,j})$ as well as $(k, b_{k,j})$ ($(c_{j,i}, i)$ as well as $(d_{j,k}, k)$) lie on π. We call π horizontally (vertically) clamped *if it is clamped in some j-th row (column). We call π* clamped *if it is horizontally or vertically clamped. We call π* clamped in cell (j, i) *if $i = k$.*

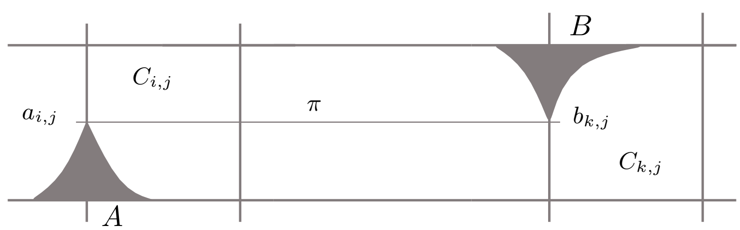

Figure 4.3: The path π is horizontally clamped in the j-th row.

Figure 4.3 shows an excerpt of F_δ and a horizontally clamped path in the j-th row of the diagram. The following Lemma from [13] subsumes our observations.

Lemma 4.8 (Each δ_F-path in $F_{\delta_F(P,Q)}$ is clamped, Alt/Godau, [13]).

1. If $\delta = \delta_F(P, Q)$, then $F_\delta(P, Q)$ contains at least one δ_F-path, and each such path is clamped.

2. If $\delta = \tilde{\delta}_F(P, Q)$, then $F_\delta(P, Q)$ contains at least one $\tilde{\delta}_F$-path, and each such path is clamped in *some cell*.

A clamped path also has a geometric interpretation. Figure 4.4 shows the geometric situations that correspond to a (horizontally) clamped path. In case (a) the reparametrization (i.e., the path) maps the point $P(i - 1)$ to the *only* point on the edge Q_j that has distance δ from $P(i-1)$. This corresponds to a $\tilde{\delta}_F$-path that is clamped in cell (i,j). In case (b) it maps the part of P between $P(i-1)$ and $P(k-1)$ to the *only* point on the edge Q_j that has distance δ from $P(i-1)$ and $P(k-1)$. This corresponds to a δ_F-path that is clamped in the j-th row. These situations cover the case of horizontally clamped paths. The geometric situations that involve a vertically clamped passage are similar, with the roles of P and Q interchanged.

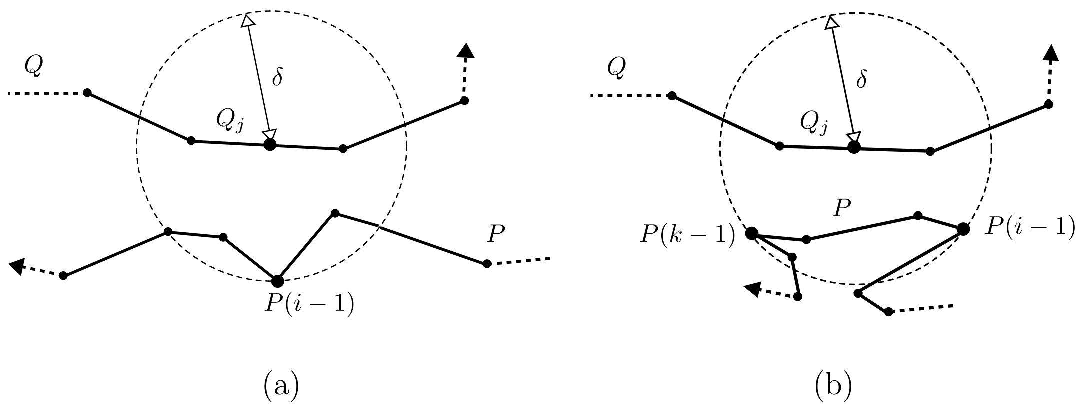

Figure 4.4: The geometric situations corresponding to a horizontally clamped path.

4.2 Minimizing the Fréchet distance

First we give a rough sketch of the basic idea of our algorithm: Assume that there is at least one translation $\tau_{\leq}$ that moves P to a Fréchet distance at most δ to Q. Then we can move P to a position $\tau_{=}$ where the Fréchet distance to Q is exactly δ. According to Lemma 4.8 the free space diagram $F_\delta(\tau_{=}(P), Q)$ then contains at least one clamped path. As a consequence, one of the geometric situations from Figure 4.4 must occur. Therefore the set of translations that attain a Fréchet distance of exactly δ is a subset of the set of translations that realize at least one of those geometric situations. The set of translations that create a geometric situation involving the two different vertices $P(i-1)$ and $P(k-1)$ from P and the edge Q_j from Q consist of two segments in transformation space, i.e., it can be described geometrically.

Now assume that the geometric situation from above is specified by the two vertices $P(i-1)$ and $P(k-1)$ and the edge Q_j. When we move P in such a way that $P(i-1)$ and $P(k-1)$ remain at distance δ from a common point on an edge of Q (i.e., we shift P 'along' Q), we will preserve one geometric situation (namely the one involving $P(i-1)$ and $P(k-1)$ and some edge of Q). During this process the Fréchet distance may vary, but at some point we will reach a placement $\tau'_{=}$ where it becomes δ again. According to Lemma 4.8 the free space diagram $F_\delta(\tau'_{=}(P), Q)$ then also contains a clamped path and another geometric situation from Figure 4.4 must occur. So whenever there is a translation that attains a Fréchet distance of exactly δ, there is a translation that realizes at least *two* such geometric situations. Moreover it can be shown that the number of translations that realize at least two such situations is only polynomial.

After this informal description of the basic ideas let us go into more detail now. Let us first take a look at the free space $F_\delta(\tau(P), Q)$ as τ varies over $\mathcal{T}_2$:

Lemma 4.9 (Continuity lemma I). *$\delta_F(\tau(P), Q)$ and $\tilde{\delta}_F(\tau(P), Q)$ are continuous functions of (the parameters of) τ.*

Proof. Let $\delta := \delta_F(P, Q)$ and consider two reparametrizations α and β of P and Q, such that $||P(\alpha(t)) - Q(\beta(t))|| \leq \delta$ for all $t \in [0, 1]$. The triangle inequality implies that $||\tau(P(\alpha(t))) - Q(\beta(t))|| \leq \delta + ||\tau||$ for all $t \in [0, 1]$, so that $\delta_F(\tau(P), Q) \leq \delta_F(P, Q) + ||\tau||$. The same argument yields that $\delta_F(P, Q) = \delta_F(\tau^{-1}(\tau(P)), Q) \leq \delta_F(\tau(P), Q) + ||\tau^{-1}|| = \delta_F(\tau(P), Q) + ||\tau||$, so that $|\delta_F(\tau(P), Q) - \delta_F(P, Q)| \leq ||\tau||$, which proves the claim. The same reasoning applies for the weak Fréchet distance. ☐

Lemma 4.10 (Continuity lemma II). *The cell boundaries $a_{i,j}$, $b_{i,j}$, $c_{i,j}$, and $d_{i,j}$ in F_δ are continuous partial functions of (the parameters of) τ. The domains of these functions are closed sets.*

Proof. Let us focus on a cell $C_{i,j}$ of the free space diagram $F_\delta(\tau(P), Q)$ as τ varies. We will show that $a_{i,j}$ and $b_{i,j}$ are continuous partial functions of (the parameters of) τ. A symmetric argument proves the claim for $c_{i,j}$, and $d_{i,j}$.

Consider the point $p = P(i-1)$ and the segment $s = Q_j$. Let p_δ denote the circle with center p and radius δ . Recall that $L_{i,j}$ denotes the left boundary of the free space of the cell $C_{i,j}$. This boundary essentially corresponds to the intersection of p_δ and s. To be more precise a point $y \in L_{i,j}$ corresponds to a point $q = Q(y) \in s$ with $||p - q|| \leq \delta$, i.e., a point in $p_\delta \cap s$, and vice versa.

Since $a_{i,j}$ and $b_{i,j}$ are defined whenever $L_{i,j}$ is non-empty we see that the domain of these two functions is the set of all translations τ such that $\tau(p_\delta)$ and s intersect, or — equivalently — such that $\tau(p)$ lies in the Minkowski sum of s with a circle of radius δ, i.e., $\text{dom}(a_{i,j}) = \text{dom}(b_{i,j}) = s \oplus (\delta \cdot \mathbb{S}^1) - p$. This is a closed set. Of course $p_\delta \cap s$ changes continuously in τ. ☐

The following result is an immediate consequence of the previous Lemma:

Corollary 4.11. *There exists $\epsilon > 0$ such that for each translation τ with $||\tau|| < \epsilon$ the following holds for all i, j, k with $i \leq k$:*

1. *If $L_{i,j} = \emptyset$, then $L_{i,j}(\tau) = \emptyset$.*
2. *If $L_{i,j} \neq \emptyset$, $L_{k,j} \neq \emptyset$, then*
   - *(a) $L_{i,j}(\tau) = \emptyset$ and $a_{i,j} = b_{i,j}$, or*
   - *(b) $L_{k,j}(\tau) = \emptyset$ and $a_{k,j} = b_{k,j}$, or*

   (c) $L_{i,j}(\tau) \neq \emptyset$, $L_{k,j}(\tau) \neq \emptyset$.

3. *If* $L_{i,j} \neq \emptyset$, $L_{k,j} \neq \emptyset$, $L_{i,j}(\tau) \neq \emptyset$, $L_{k,j}(\tau) \neq \emptyset$, *and*

   (a) $a_{i,j} > b_{k,j}$, *then* $a_{i,j}(\tau) > b_{k,j}(\tau)$.

   (b) $a_{i,j} < b_{k,j}$, *then* $a_{i,j}(\tau) < b_{k,j}(\tau)$.

The corresponding statement holds for $c_{i,j}$, $d_{k,j}$, $B_{i,j}$, *and* $B_{k,j}$.

Definition 4.12 (Configuration). *A triple* $c = (p, p', s)$ *that consists of two (not necessarily distinct) vertices* p *and* p' *of* Q (P) *and an edge* s *of* P (Q) *is called an* h-configuration (v-configuration) *of* P *and* Q. *A* configuration *is an h-configuration or a v-configuration.*

If $p = p'$ *then* c *is called* degenerate. *Otherwise it is called* non-degenerate.

Let τ *be a translation. If* c *is an h-configuration, then* $\tau(c) := (p, p', \tau(s))$, *whereas if* c *is a v-configuration, then* $\tau(c) := (\tau(p), \tau(p'), s)$.

Let us briefly rephrase the statement of Lemma 4.8 and the discussion following it:

Observation 4.13.

1. *If* $\delta = \delta_F(P, Q)$, *then there is a configuration* $c = (p, p', s)$ *of* P *and* Q *such that there exist* at most two *points* $q \in s$ *with* $||p - q|| = ||p' - q|| = \delta$.

2. *If* $\delta = \tilde{\delta}_F(P, Q)$, *then there is a* degenerate *configuration* $c = (p, p, s)$ *of* P *and* Q *such that there exist* at most two *points* $q \in s$ *with* $||p - q|| = \delta$.

Definition 4.14 (δ–critical translation). *A translation* τ *is called* δ–critical for a configuration c *of* P *and* Q, *if* $\tau(c) = (p, p', s)$ *and there exist* at most two *points* $q \in s$ *s.th.* $||p - q|| = ||p' - q|| = \delta$. *A translation is called* δ-critical *if it is* δ*-critical for some configuration. The set of all translations that are* δ*–critical for* c *will be denoted by* $\mathcal{T}^{\delta}_{crit}(c)$.

Observation 4.15. *Let* $c = (P(i-1), P(k-1), Q_j)$, *with* $i \leq k$, *be an h-configuration. Then* $\mathcal{T}^{\delta}_{crit}(c) \subseteq \mathrm{dom}(a_{i,j}) \cap \mathrm{dom}(b_{k,j})$ *and* $a_{i,j}(\tau) = b_{k,j}(\tau)$ *for all* $\tau \in \mathcal{T}^{\delta}_{crit}(c)$.

Of course the statement of this observation remains true if we consider v-configurations (where the roles of P and Q are interchanged).

Lemma 4.16 (Characteristic feasible translations - weak version). *If there is a translation* $\tau_{\leq}$ *such that* $\delta_F(\tau_{\leq}(P), Q) \leq \delta$ $(\tilde{\delta}_F(\tau_{\leq}(P), Q) \leq \delta)$ *then there is a translation* $\tau_{=}$ *that is* δ*–critical such that* $\delta_F(\tau_{=}(P), Q) = \delta$ $(\tilde{\delta}_F(\tau_{=}(P), Q) = \delta)$.

Proof. We formulate the proof in terms of the Fréchet distance only. The proof for the case of the weak Fréchet distance can be copied verbatim. Pick any translation $\tau_{>}$ such that $\delta_F(\tau_{>}(P), Q) > \delta$. Since $\delta_F(\tau(P), Q)$ is a continuous function of τ according to Lemma 4.9, there exists a translation $\tau_{=}$ on any curve between $\tau_{\leq}$ and $\tau_{>}$ in translation space such that $\delta_F(\tau_{=}(P), Q) = \delta$. By Observation 4.13 the translation $\tau_{=}$ is critical for some configuration. □

This result states that in order to check if there is a translation that moves P into Fréchet distance at most δ to Q, it is sufficient to check all the δ–critical translations.

So let us take a closer look at the set of δ–critical translations in $\mathcal{T}_2$: Let $c = (P(i-1), P(k-1), Q_j)$, with $i \leq k$, be an h-configuration. Consider the points $p = P(i-1)$, $p' = P(k-1)$ and the segment $s = Q_j$.

First assume that $i < k$, i.e., c is non-degenerate. By definition a translation τ is δ–critical for c iff there exist at most two points $q \in s$ s.th. $||\tau(p) - q|| = ||\tau(p') - q|| = \delta$, i.e., one of the two points $\tau(p_\delta) \cap \tau(p'_\delta) = \tau(p_\delta \cap p'_\delta)$ lies on s. So if $p_\delta \cap p'_\delta = \{r, r'\}$ we have that $\mathcal{T}^\delta_{crit}(c) = \{\tau \mid \tau(r) \in s \text{ or } \tau(r') \in s\} = (s - r) \cup (s - r')$. Thus for a non-degenerate configuration (which corresponds to case (b) in Figure 4.4) the set of δ–critical translations is described by two parallel line segments in translation space, where each line segment is a translate of s.

In case that $i = k$, i.e., c is degenerate, a translation τ is δ–critical for c iff $\mathrm{d}(\tau(p), s) = \delta$, i.e., the circle $\tau(p_\delta)$ 'touches' s, or — equivalently — such that $\tau(p)$ lies on the boundary of the Minkowski sum of s with a circle of radius δ, i.e., $\mathcal{T}^\delta_{crit}(c) = \mathsf{bd}(s \oplus (\delta \cdot \mathbb{S}^1)) - p = \mathsf{bd}(\mathrm{dom}(a_{i,j}))$. Thus, for a degenerate configuration (which is case (a) in Figure 4.4) the set of δ–critical translations is described by a 'racetrack' in translation space, which is the locus of all points having distance δ to a translate of s. Such a 'racetrack' consists of two parallel line segments in translation space, where each line segment is a translate of s and two semicircles of radius δ.

Again a similar statement remains true if we consider v-configurations (where the roles of P and Q are interchanged). These results are subsumed in the following

Observation 4.17. *For a non-degenerate configuration involving the segment s, the set of δ–critical translations is described by two parallel line segments in translation space, where each line segment is a translate of s. For a degenerate configuration involving the segment s, the set of δ–critical translations consists of two parallel line segments in translation space, where each line segment is a translate of s and two semicircles of radius δ. The critical translations of two different configurations either overlap or intersect at most four times.*

Lemma 4.18 (Characteristic feasible translations for δ_F - strong version). *Let c be a configuration and assume that there is a translation $\tau_= \in \mathcal{T}^\delta_{crit}(c)$ s.th. $\delta_F(\tau_=(P), Q) = \delta$. Let $\mathbf{s}$ be the segment or semicircle of $\mathcal{T}^\delta_{crit}(c)$ containing $\tau_=$ and assume that for the two endpoints τ_1, τ_2 of $\mathbf{s}$ we have that $\delta_F(\tau_1(P), Q) > \delta$ and $\delta_F(\tau_2(P), Q) > \delta$.*

Then there is a configuration $c' \neq c$ with $|\mathcal{T}^\delta_{crit}(c) \cap \mathcal{T}^\delta_{crit}(c')| \leq 4$, and a translation $\tau'_= \in \mathcal{T}^\delta_{crit}(c) \cap \mathcal{T}^\delta_{crit}(c')$, such that $\delta_F(\tau'_=(P), Q) = \delta$.

Proof. Let us assume wlog that $c = (P(i-1), P(k-1), Q_j)$, with $i \leq k$ is an h-configuration. The case that c is a v-configuration is analogous.

It follows from Lemma 4.9 that the set $\{\tau \mid \delta_F(\tau(P), Q) \leq \delta\}$ is a closed set. Therefore the set $\mathbf{s}_\leq := \{\tau \in \mathbf{s} \mid \delta_F(\tau(P), Q) \leq \delta\}$ consists of segments or circular arcs on $\mathbf{s}$ whose endpoints also belong to $\mathbf{s}_\leq$ (loosely speaking, $\mathbf{s}_\leq$ is a closed subset of $\mathbf{s}$). Let $\mathbf{s}^0_\leq$ denote the set of these endpoints; this set is non-empty since $\tau_= \in \mathbf{s}_\leq$. Lemma 4.9 also implies that for any $\tau \in \mathbf{s}^0_\leq$ we have that $\delta_F(\tau(P), Q) = \delta$. Pick any $\tau'_= \in \mathbf{s}^0_\leq$.

Since $\{\tau_1, \tau_2\} \cap \mathbf{s}_\leq = \emptyset$ there is an $\epsilon' > 0$ such that for all $0 < \mu < \epsilon'$ there is a $\tau_> \in \mathbf{s} \setminus \mathbf{s}_\leq$ with $||\tau'_= - \tau_>|| \leq \mu$.

We have for all $\tau \in \mathbf{s}$ that (c.f. Observation 4.15)

$$L_{i,j}(\tau) \neq \emptyset, L_{k,j}(\tau) \neq \emptyset, \text{ and } a_{i,j}(\tau) = b_{k,j}(\tau). \tag{4.1}$$

Let $\epsilon > 0$ be such that a translation that is at most ϵ away from $\tau'_=$ does not change the relative position of all spikes that are not aligned. Such an ϵ indeed exists according to Corollary 4.11.

We can conclude that for all $0 < \mu < \mathsf{min}(\epsilon, \epsilon')$ there exists some $\tau_> \in \mathbf{s} \setminus \mathbf{s}_\leq$ with $||\tau'_= - \tau_>|| \leq \mu$ that does not change the relative position of any spikes that are not aligned (in the sense of Corollary 4.11).

Since $F_\delta(\tau_>(P), Q)$ does not contain any δ_F-paths, the diagram must differ from $F_\delta(\tau'_=(P), Q)$. The only possibilities that prevent such paths are

1. $L_{r,t}(\tau'_=) \neq \emptyset$, but $L_{r,t}(\tau_>) = \emptyset$ for some r, t, or
2. $a_{r,s}(\tau'_=) = b_{t,s}(\tau'_=)$, but $a_{r,s}(\tau_>) > b_{t,s}(\tau_>)$ for some r, s, t.
3. $B_{r,t}(\tau'_=) \neq \emptyset$, but $B_{r,t}(\tau_>) = \emptyset$ for some r, t, or
4. $c_{r,s}(\tau'_=) = d_{t,s}(\tau'_=)$, but $c_{r,s}(\tau_>) > d_{t,s}(\tau_>)$ for some r, s, t.

In the first case Corollary 4.11 yields that $a_{r,t}(\tau'_=) = b_{r,t}(\tau'_=)$ and our considerations (4.1) above show that $(r, t) \neq (i, j)$ and $(r, t) \neq (k, j)$. Thus we get a new h-configuration $c' = (P(r-1), P(r-1), Q_t)$ for which $\tau'_=$ is critical.

In the second case our considerations (4.1) above show that $(r, s, t) \neq (i, j, k)$. Thus we get a new h-configuration $c' = (P(r-1), P(s-1), Q_t)$ for which $\tau'_=$ is critical.

In any case $\tau_> \notin \mathcal{T}^\delta_{crit}(c')$, so in the neighbourhood of $\tau'_=$ the two sets $\mathcal{T}^\delta_{crit}(c)$ and $\mathcal{T}^\delta_{crit}(c')$ do not overlap. Since $\tau'_= \in \mathcal{T}^\delta_{crit}(c) \cap \mathcal{T}^\delta_{crit}(c')$ we can conclude that they do not overlap at all.

The other two cases are analogous and yield a v-configuration c' for which $\tau'_=$ is critical.

□

The proof of this Lemma is easily modified to yield the corresponding result for the weak Fréchet distance; in fact only the cases 1. and 3. need to be taken into account. This yields:

Lemma 4.19 (Characteristic feasible translations for $\tilde{\delta}_F$ - strong version). *Let c be a degenerate configuration and assume that there is a translation $\tau_= \in \mathcal{T}^\delta_{crit}(c)$ s.th. $\tilde{\delta}_F(\tau_=(P), Q) = \delta$. Let $\mathbf{s}$ be the segment or semicircle of $\mathcal{T}^\delta_{crit}(c)$ containing $\tau_=$ and assume that for the two endpoints τ_1, τ_2 of $\mathbf{s}$ we have that $\tilde{\delta}_F(\tau_1(P), Q) > \delta$ and $\tilde{\delta}_F(\tau_2(P), Q) > \delta$.*

Then there is a degenerate configuration $c' \neq c$ with $|\mathcal{T}^\delta_{crit}(c) \cap \mathcal{T}^\delta_{crit}(c')| \leq 4$, and a translation $\tau'_= \in \mathcal{T}^\delta_{crit}(c) \cap \mathcal{T}^\delta_{crit}(c')$, such that $\tilde{\delta}_F(\tau'_=(P), Q) = \delta$.

The previous two results show that the translations that correspond to the endpoints of **s** deserve special attention:

Definition 4.20 (δ–supercritical translation). *Let c be a configuration and assume that $\mathcal{T}_{crit}^{\delta}(c)$ is the disjoint union of the sets h, h', s, and s' where s and s' are parallel line segments, and h and h' are either empty (if c is non-degenerate) or semicircles of radius δ (if c is degenerate). The translations that correspond to the endpoints of s and s' will be called δ–*supercritical*. By ${\mathcal{T}'_{crit}}^{\delta}(c)$ we denote the set of the four translations that are δ–supercritical for c.*

Lemma 4.21 below summarizes our considerations: In order to decide whether there is a translation τ s.th. $\delta_F(\tau(P), Q) \leq \delta$, we can restrict our attention to translations that are δ–critical for two *different* configurations or that are δ–supercritical.

Lemma 4.21 (Characteristic feasible translations for δ_F). *If there is a translation $\tau_{\leq}$ such that $\delta_F(\tau_{\leq}(P), Q) \leq \delta$ then*

1. *either there is a configuration c and a translation $\tau'_{\leq} \in {\mathcal{T}'_{crit}}^{\delta}(c)$, such that $\delta_F(\tau'_{\leq}(P), Q) \leq \delta$,*
2. *or there are two different configurations $c \neq c'$ with $|\mathcal{T}_{crit}^{\delta}(c) \cap \mathcal{T}_{crit}^{\delta}(c')| \leq 4$, and a translation $\tau'_{=} \in \mathcal{T}_{crit}^{\delta}(c) \cap \mathcal{T}_{crit}^{\delta}(c')$, such that $\delta_F(\tau'_{=}(P), Q) = \delta$.*

Proof. By Lemma 4.16 there is a translation $\tau_{=}$ with $\delta_F(\tau_{=}(P), Q) = \delta$ that is δ–critical for some configuration c, i.e., $\tau_{=} \in \mathcal{T}_{crit}^{\delta}(c)$. Let **s** be the segment or semicircle of $\mathcal{T}_{crit}^{\delta}(c)$ containing $\tau_{=}$ and let $\tau_1, \tau_2 \in {\mathcal{T}'_{crit}}^{\delta}(c)$ be the two endpoints of **s**.

1. If $\delta_F(\tau_1(P), Q) \leq \delta$ or $\delta_F(\tau_2(P), Q) \leq \delta$ the claim of part 1. follows with $\tau'_{\leq} := \tau_1$ or $\tau'_{\leq} := \tau_2$, respectively.
2. If $\delta_F(\tau_1(P), Q) > \delta$ and $\delta_F(\tau_2(P), Q) > \delta$ the claim of part 2. follows from Lemma 4.18.

□

Again, the proof of this Lemma is easily modified to yield the corresponding result for the weak Fréchet distance:

Lemma 4.22 (Characteristic feasible translations for $\tilde{\delta}_F$). *If there is a translation $\tau_{\leq}$ such that $\tilde{\delta}_F(\tau_{\leq}(P), Q) \leq \delta$ then*

1. *either there is a degenerate configuration c and a translation $\tau'_{\leq} \in {\mathcal{T}'_{crit}}^{\delta}(c)$, such that $\tilde{\delta}_F(\tau'_{\leq}(P), Q) \leq \delta$,*
2. *or there are two different degenerate configurations $c \neq c'$ with $|\mathcal{T}_{crit}^{\delta}(c) \cap \mathcal{T}_{crit}^{\delta}(c')| \leq 4$, and a translation $\tau'_{=} \in \mathcal{T}_{crit}^{\delta}(c) \cap \mathcal{T}_{crit}^{\delta}(c')$, such that $\tilde{\delta}_F(\tau'_{=}(P), Q) = \delta$.*

So in order to solve the decision problem for a given δ it is sufficient to check all translations τ that correspond either to δ–supercritical translations or to the intersection of the set of non-overlapping δ–critical translations of two *distinct* configurations.

Algorithm *Fréchet_Match*(P, Q, δ)
Input: Two polygonal curves $P, Q \in \mathcal{K}^0$, and $\delta \geq 0$.
Output: Decides whether there exists a translation τ such that $\delta_F(\tau(P), Q) \leq \delta$.

for all configurations c
  do Compute τ'_{crit}, the δ–supercritical translations for c
    for $\tau \in \tau'_{crit}$
      do if $\delta_F(\tau(P), Q) \leq \delta$
        then return (TRUE)
for all pairs of configurations c and c'
  do Compute τ_{crit}, the δ–critical translations for c
    Compute τ'_{crit}, the δ–critical translations for c'
    for $\tau \in \tau_{crit} \cap \tau'_{crit}$
      do if $\delta_F(\tau(P), Q) \leq \delta$
        then return (TRUE)
return (FALSE)

The correctness of Algorithm *Fréchet_Match* follows from Lemma 4.21. The complexity of the algorithm depends on the number of configurations. There are $\mathcal{O}(m^2 n)$ many h-configurations and $\mathcal{O}(n^2 m)$ many v-configurations. We thus have altogether $\mathcal{O}\big((mn)^2(m+n)^2\big)$ translations for each of which we check in $\mathcal{O}(mn)$ time if it brings P into Fréchet distance at most δ to Q. This solves Problem 4.2 for the Fréchet distance and yields the following Theorem:

Theorem 4.23 (Correctness and complexity of *Fréchet_Match***).** *In* $\mathcal{O}\big((mn)^3(m+n)^2\big)$ *time Algorithm Fréchet_Match decides, whether there is a translation* $\tau \in \mathcal{T}_2$ *such that* $\delta_F(\tau(P), Q) \leq \delta$.

We can copy Algorithm *Fréchet_Match* almost verbatim to devise a procedure for solving the corresponding decision problem for the weak Fréchet distance.

Algorithm *Weak_Fréchet_Match*(P, Q, δ)
Input: Two polygonal curves $P, Q \in \mathcal{K}^0$, and $\delta \geq 0$.
Output: Decides whether there exists a translation τ such that $\tilde{\delta}_F(\tau(P), Q) \leq \delta$.

for all degenerate configurations c
  do Compute τ'_{crit}, the δ–supercritical translations for c
    for $\tau \in \tau'_{crit}$
      do if $\tilde{\delta}_F(\tau(P), Q) \leq \delta$
        then return (TRUE)
for all pairs of degenerate configurations c and c'
  do Compute τ_{crit}, the δ–critical translations for c

Compute τ'_{crit}, the δ–critical translations for c'
for $\tau \in \tau_{crit} \cap \tau'_{crit}$
  do if $\tilde{\delta}_F(\tau(P), Q) \leq \delta$
    then return (TRUE)
return (FALSE)

The correctness of Algorithm *Weak_Fréchet_Match* also follows from Lemma 4.22. Since there are only $\mathcal{O}(mn)$ many degenerate configurations, we altogether have $\mathcal{O}\big((mn)^2\big)$ translations for each of which we check in $\mathcal{O}(mn)$ time if it brings P into weak Fréchet distance at most δ to Q.

Theorem 4.24 (Correctness and complexity of *Weak_Fréchet_Match***).** *The algorithm Weak_Fréchet_Match decides in $\mathcal{O}\big((mn)^3\big)$ time whether there is a translation $\tau \in \mathcal{T}_2$ such that $\tilde{\delta}_F(\tau(P), Q) \leq \delta$.*

In order to find a translation that minimizes the (weak) Fréchet distance between the two polygonal curves one can apply the parametric search paradigm of Megiddo [40] and the improvement by Cole [29]. Details can be found in [17] and [50]. The latter also gives a generalization of our technique to other sets of transformations, like, e.g., rigid motions.

4.3 Approximately minimizing the Fréchet distance

The algorithms we described so far cannot be considered to be efficient. To remedy this situation, we present approximation algorithms which do not necessarily compute the optimal transformation, but one that yields a (weak) Fréchet distance which differs from the optimum value by a constant factor only. To this end, we generalize the notion of a *reference point*, c.f. [9] and [8], to the Fréchet metrics and observe that all reference points for the Hausdorff distance are also reference points for the (weak) Fréchet distance.

We first need the concept of a *reference point* that was introduced in [8]. A reference point of a figure is a characteristic point with the property that similar figures have reference points that are close to each other. Therefore we get a reasonable matching of two figures if we simply align their reference points.

Definition 4.25 (Reference point, [8]). *Let $\mathcal{K}$ be a set of planar figures and $\delta : \mathcal{K} \to \mathbb{R}$ be a distance measure on $\mathcal{K}$. A mapping $\mathrm{R} : \mathcal{K} \to \mathbb{R}^2$ is called a δ–*reference point *for $\mathcal{K}$ of* quality $c > 0$ *with respect to a set of transformations $\mathcal{T}$ on $\mathcal{K}$, if the following holds for any two figures $P, Q \in \mathcal{K}$ and each transformation $\tau \in \mathcal{T}$:*

$$\text{(Equivariance)} \qquad \mathrm{R}(\tau(P)) = \tau(\mathrm{R}(P)) \tag{4.2}$$

$$\text{(Lipschitz continuity)} \qquad ||\,\mathrm{R}(P) - \mathrm{R}(Q)|| \leq c \cdot \delta(P, Q). \tag{4.3}$$

The set of δ–reference points for $\mathcal{K}$ of quality c with respect to $\mathcal{T}$ is denoted $\mathcal{R}(\mathcal{K}, \delta, c, \mathcal{T})$.

So if δ is a metric on $\mathcal{K}$, a reference point is a Lipschitz-continuous mapping between the metric spaces $(\mathcal{K}, \delta)$ and $(\mathbb{R}^2, ||\cdot||)$ with Lipschitz constant c, which is equivariant under $\mathcal{T}$. Various reference points are known for a variety of distance measures and classes of transformations, like, e.g., the centroid of a convex polygon which is a reference point of quality 11/3 for translations, using the area of the symmetric difference as a distance measure, see [12]. However, most work on reference points has focused on the Hausdorff distance, see [8].

We will only mention the following result that provides a δ_H–reference point for polygonal curves with respect to similarities, the so called *Steiner point*. The Steiner point of a polygonal curve is the weighted average of the vertices of the convex hull of the curve, where each vertex is weighted by its exterior angle divided by 2π.

Theorem 4.26 (Steiner point, Aichholzer et al., [8]). The Steiner point is a δ_H–reference point with respect to similarities of quality $4/\pi$. It can be computed in linear time.

Note that the Steiner point is an *optimal* δ_H–reference point with respect to similarities, i.e., the quality of any δ_H–reference point for that transformation class is at least $4/\pi$, see [8].

Two feasible reparametrizations α and β of P and Q demonstrate, that for each point $P(\alpha(t))$ there is a point $Q(\beta(t))$ with $||P(\alpha(t)) - Q(\beta(t))|| \leq \delta$ (and vice versa), thus $\delta_H(P,Q) \leq \tilde{\delta}_F(P,Q) \leq \delta_F(P,Q)$. This is summarized in the following:

Observation 4.27. *Let $c > 0$ be a constant and $\mathcal{T}$ be a set of transformations on $\mathcal{K}^0 \cup \mathcal{K}^1$. Then each δ_H–reference point with respect to $\mathcal{T}$ is also a $\tilde{\delta}_F$–reference point with respect to $\mathcal{T}$, and each $\tilde{\delta}_F$–reference point is also a δ_F–reference point of the same quality, i.e.,*

$$\mathcal{R}(\mathcal{K}^0 \cup \mathcal{K}^1, \delta_H, c, \mathcal{T}) \subseteq \mathcal{R}(\mathcal{K}^0 \cup \mathcal{K}^1, \tilde{\delta}_F, c, \mathcal{T}) \subseteq \mathcal{R}(\mathcal{K}^0 \cup \mathcal{K}^1, \delta_F, c, \mathcal{T}).$$

This shows that we can use the known δ_H–reference points to obtain $\tilde{\delta}_F$–reference points and δ_F–reference points. However, since for *non-closed* curves each reparametrization has to map $P(0)$ to $Q(0)$, the distance $||P(0) - Q(0)||$ is a lower bound for $\tilde{\delta}_F(P,Q)$ as well as $\delta_F(P,Q)$. So we get a new reference point that is substantially better than all known reference points for the Hausdorff distance.

Observation 4.28. *The mapping*

$$\mathrm{R}_o : \begin{cases} \mathcal{K}^0 \to \mathbb{R}^2 \\ P \mapsto P(0) \end{cases}$$

is a $\tilde{\delta}_F$–reference point for curves of quality 1 *with respect to translations, i.e.,*

$$\mathrm{R}_o \in \mathcal{R}(\mathcal{K}^0, \tilde{\delta}_F, 1, \mathcal{T}_2).$$

The quality of this reference point, i.e., 1, is better than the quality of the Steiner point, which is $4/\pi$. Since the latter is an optimal reference point for the Hausdorff distance, this shows that for the Fréchet distance substantially better reference points exist. For closed curves however, R_o is not defined at all, so we have to stick to the reference points we get from Observation 4.27.

Based on the existence of these reference points for $\mathcal{T}_2$ we obtain the following algorithm for approximate matchings of (closed) curves with respect to the (weak) Fréchet distance under the group of translations, which is the same procedure as already used in [8] for the Hausdorff distance. In what follows, $\mathcal{C} \in \{\mathcal{K}^0, \mathcal{K}^1\}$, and $\delta \in \{\tilde{\delta}_F, \delta_F\}$.

Algorithm *Approx_Fréchet_Match*(P, Q, R)
Input: Two polygonal curves $P, Q \in \mathcal{C}$ and a reference point $\mathrm{R} \in \mathcal{R}(\mathcal{C}, \delta, c, \mathcal{T}_2)$.
Output: A value δ_{approx} and a translation τ_{approx} such that

1. $\delta_{approx} = \delta(\tau_{approx}(P), Q)$, and
2. $\delta_{approx} \leq (c+1) \cdot \delta_{opt}$, where $\delta_{opt} = \mathsf{min}_\tau \delta(\tau(P), Q)$.

1. Compute $\mathrm{R}(P)$ and $\mathrm{R}(Q)$
2. $\tau_{approx} := \mathrm{R}(Q) - \mathrm{R}(P)$
3. $\delta_{approx} := \delta(\tau_{approx}(P), Q)$
4. **return** $(\delta_{approx}, \tau_{approx})$

Theorem 4.29 (Correctness and complexity of *Approx_Fréchet_Match***).** *Suppose that* R *can be computed in* $\mathcal{O}(T_{\mathrm{R}}(n))$ *time. Then Algorithm Approx_Fréchet_Match produces a* $(c+1)$*-approximation to Problem 4.1 in* $\mathcal{O}\big(mn\log^{k+1}(mn) + T_{\mathrm{R}}(m) + T_{\mathrm{R}}(n)\big)$ *time, where* $k = 0$ *if* $\delta = \delta_F$*, and* $k = 1$ *if* $\delta = \tilde{\delta}_F$*.*

Proof. The correctness of this procedure follows easily from the results of [8]. The proof of the claimed time bounds follows from the results (Theorems 6, 7b, and 11) of [13]. □

Note that with an idea from [45] it is possible to reduce the approximation constant for reference point based matching to $(1+\epsilon)$ for any $\epsilon > 0$; the idea places a sufficiently small grid of size $\mathcal{O}(1/\epsilon^2)$ around the reference point of Q and checks each grid point as a potential image point for the reference point of P. The runtime increases by a factor proportional to the grid size.

4.3.1 Reference points for affine maps

We conclude this chapter with a negative result on the approximate matching problem. We have already seen that – compared to the Hausdorff distance – better reference points for the Fréchet distance exist (for polygonal curves), i.e.,

$$\mathcal{R}(\mathcal{K}^0, \delta_H, \mathcal{T}_2) \subsetneq \mathcal{R}(\mathcal{K}^0, \tilde{\delta}_F, \mathcal{T}_2),$$

where $\mathcal{R}(\mathcal{K}^0, \delta, \mathcal{T}_2) := \bigcup_{c>0} \mathcal{R}(\mathcal{K}^0, \delta, c, \mathcal{T}_2)$ is the set of all δ–reference points for $\mathcal{K}^0$ with respect to $\mathcal{T}_2$.

Note that we can easily strengthen Observation 4.28 as follows:

Observation 4.30. *Let* R_o *be as above, and* $\mathcal{T}$ *be a set of transformations on* $\mathcal{K}^0$*, such that* R_o *is equivariant under* $\mathcal{T}$*. Then* R_o *is a* $\tilde{\delta}_F$*–reference point for* $\mathcal{K}^0$ *of quality* 1 *with respect to* $\mathcal{T}$*, i.e.,*

$$\mathrm{R}_o \in \mathcal{R}(\mathcal{K}^0, \tilde{\delta}_F, 1, \mathcal{T}).$$

In particular this implies that $\mathcal{R}(\mathcal{K}^0, \tilde{\delta}_F, \mathcal{A}_2) \neq \emptyset$, where $\mathcal{A}_2$ is the set of affine transformations of the plane. This is complemented by the following result:

Theorem 4.31 (Non-existence of δ_H-reference points for affine maps). *There are no* δ_H*-reference points for affine maps, i.e.,*

$$\mathcal{R}(\mathcal{K}^0 \cup \mathcal{K}^1, \delta_H, \mathcal{A}_2) = \emptyset.$$

Proof. First observe that for two curves $P_1, P_2 \in \mathcal{K}^0 \cup \mathcal{K}^1$ with $P_1 \neq P_2$ and $\delta_H(P_1, P_2) = 0$, and a reference point $\mathrm{R} \in \mathcal{R}(\mathcal{K}^0 \cup \mathcal{K}^1, \delta_H, \mathcal{T})$, we have that $||\,\mathrm{R}(P_1) - \mathrm{R}(P_2)|| = 0$, so $\mathrm{R}(P_1) = \mathrm{R}(P_2)$; this means that δ_H–reference points only take the geometry of the curves into account.

In the following $c(X)$ denotes the center of gravity of the vertices of the convex hull of X. Assume R is a δ_H–reference point with respect to affine maps.

Let Δ be a curve in the shape of an equilateral triangle. Let α_r be the affine map that rotates Δ around $c(\Delta)$ by an angle of $2\pi/3$ in counterclockwise order. Let $\Delta_r = \alpha_r(\Delta)$ denote the image of Δ under α_r. Note that $c(\Delta)$ is the only fixpoint of α_r. Now since R is a δ_H–reference point, and $\delta_H(\Delta, \Delta_r) = 0$, we can conclude from our initial remark that

$$\mathrm{R}(\Delta) = \mathrm{R}(\Delta_r) = \mathrm{R}(\alpha_r(\Delta)) = \alpha_r(\mathrm{R}(\Delta)),$$

so $\mathrm{R}(\Delta)$ is a fixpoint of α_r, and thus $\mathrm{R}(\Delta) = c(\Delta)$. Now consider an arbitrary triangle Δ'. Since Δ' is the image of an equilateral triangle Δ under some affine transformation α',

$$\mathrm{R}(\Delta') = \mathrm{R}(\alpha'(\Delta)) = \alpha'(\mathrm{R}(\Delta)) = \alpha'(c(\Delta)).$$

Since c is invariant under affine maps,

$$\mathrm{R}(\Delta') = c(\alpha'(\Delta)) = c(\Delta').$$

However, as Figure 4.5 illustrates, c is not a δ_H–reference point, since it is not Lipschitz-continuous. To this end, observe that

$$\lim_{\substack{h\to 0\\ w\to\infty}} \delta_H(\Delta_1, \Delta_2) = 0, \quad \text{whereas} \quad \lim_{\substack{h\to 0\\ w\to\infty}} ||c(\Delta_1) - c(\Delta_2)|| = \infty.$$

□

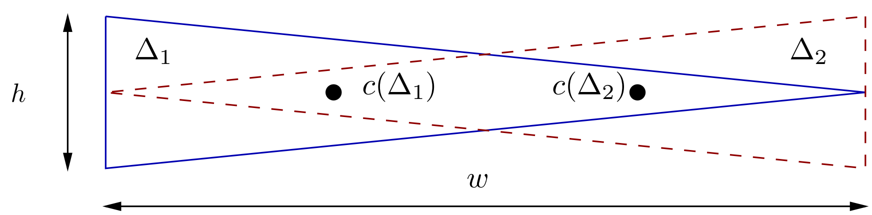

Figure 4.5: The center of gravity of the vertices of the convex hull is not a δ_H–reference point.