Chapter 5

Matching Curves with respect to the Fréchet Distance

In this chapter we consider the problem of matching two curves in d-dimensional space with respect to the Fréchet distance. As transformation classes we consider subsets of affine transformations which can be parameterized by a *rational parameterization*. We will give a formal definition later in Definition 25, but to get a first impression, all we need to know is that this notion captures many commonly used transformation classes such as translations and arbitrary affine transformations in d dimensions, and rotations, rigid motions, and similarities in two and three dimensions. In general, each affine transformation in $\mathbb{R}^d$ is composed of a linear transformation and a translation, and can thus be described* by a pair (A, t) where A is a $d \times d$ matrix representing the linear transformation and t is a d-dimensional translation. For an affine transformation $(A, t) \in \mathcal{T}$ and a polygonal curve $P : [0, m] \to \mathbb{R}^d$ we denote by $AP + t$ the curve P transformed by (A, t), i.e., $AP + t : [0, m] \to \mathbb{R}^d; s \mapsto AP(s) + t$. For the special set of d-dimensional translations we use the notation $P + t$ to denote the translation of P by the vector t, i.e., $P + t : [0, m] \to \mathbb{R}^d; s \mapsto P(s) + t$. In this chapter we consider the following matching problem:

Problem 6 (Fréchet Optimization) *Given two polygonal curves P, Q in $\mathbb{R}^d$ and a rationally parameterized subset $\mathcal{T}$ of the affine transformations in $\mathbb{R}^d$. Compute $\min_{(A,t)\in\mathcal{T}} \delta_{\mathrm{F}}(AP + t, Q)$ together with a transformation in $\mathcal{T}$ that minimizes the Fréchet distance between P and Q.*

Problem 6 is an *optimization problem*, since it asks for the transformation that minimizes the Fréchet distance. We solve this problem by first attacking the decision variant of it:

Problem 7 (Fréchet Decision) *Given two polygonal curves P, Q in $\mathbb{R}^d$, a rationally represented subset $\mathcal{T}$ of the affine transformations in $\mathbb{R}^d$, and a parameter $\varepsilon > 0$. Decide whether there exists a transformation $(A, t) \in \mathcal{T}$ such that $\delta_{\mathrm{F}}(AP + t, Q) \leq \varepsilon$, and if so compute such a transformation.*

We first solve the decision problem for a given $\varepsilon > 0$, and then solve the optimization problem by applying Megiddo's parametric search technique [62] to find the optimal ε.

*Note that the use of homogenous coordinates would simplify the notation, however we chose to stick to Cartesian coordinates in order to keep a consistent notation throughout this thesis.

In Section 5.2 we present an approach to solve both Problem 6 and Problem 7 for the case of translations in $\mathbb{R}^d$. We will see that this approach generalizes to the more general transformation classes of rationally parameterized affine transformations, which we present in Section 5.3. The reader might find it strange that we did not state the results for translations as a special case of the results for rationally parameterized affine transformations. We present the easier case of translations first in order to convey the general concepts in a more intuitive setting. Then it is rather easy to see the generalization to rationally parameterized affine transformations. Note that the solution for the case of two-dimensional curves and two-dimensional translations has been established in collaboration with Helmut Alt and Christian Knauer [17]. We utilize the basic ingredients in this thesis. However the approach we present is formulated in a completely different way (using configuration spaces) in order to support its generalization to higher dimensions. The parametric search to optimize Problem 7 is completely original in this thesis.

Table 5.1: Summary of the results of Chapter 5 of finding a transformation which minimizes the Fréchet distance between two polygonal curves in d dimensions, of complexity m and n; $N := m + n$. *DOF* denotes the number of *degrees of freedom* of the rational parameterization of the transformation class, see Definition 25. The results below the line are special cases of the result for d-dimensional transformations with k degrees of freedom.

Transformations	Dim.	DOF	Time	Space
Translations	d	d	$O(N^{3d+2} \log N)$	$O(N^{3d})$
Transformations	d	k	$O(N^{3k+2} \log N)$	$O(N^{3k})$
Affine Transf.	d	$d^2 + d$	$O(N^{3(d^2+d)+2} \log N)$	$O(N^{3(d^2+d)})$
Scalings	d	1	$O(N^5 \log N)$	$O(N^3)$
Rotations	2	1	$O(N^5 \log N)$	$O(N^3)$
Rotations	3	3	$O(N^{11} \log N)$	$O(N^9)$
Rigid Motions	2	3	$O(N^{11} \log N)$	$O(N^9)$
Rigid Motions	3	6	$O(N^{20} \log N)$	$O(N^{18})$
Similarities	2	4	$O(N^{14} \log N)$	$O(N^{12})$
Similarities	3	7	$O(N^{23} \log N)$	$O(N^{21})$

We show in Section 5.2 and Section 5.3 that for rationally parameterized subsets of the affine transformations with k degrees of freedom Problem 6 can be solved, i.e., the transformation can be found that minimizes the Fréchet distance under this class of transformations, in $O(N^{3k+2} \log N)$ time, where N is the sum of the complexities of both input curves. For d-dimensional translations this becomes $O(N^{3d+2} \log N)$ time, see Table 5.1 for a summary of the result applied to different standard classes of transformations. In Section 6.3 we consider variants of Problem 6 which can be solved using the same techniques as in Section 5.2 and Section 5.3. In particular, we show that we can solve Problem 6 for the weak Fréchet distance in $O(N^{2k+2} \log N)$ time, for closed polygonal curves in $O(N^{2k+2} \log^2 N)$ time, and we can solve a partial matching variant in $O(N^{2k+2} \log^2 N)$ time.

Note that the time and space complexity of the algorithms are polynomial, yet very large. However they are the first algorithms for higher dimensions and non-trivial transformation classes. The only lower bound on the complexity of the configuration space that we consider to compute curves under translations in dimensions $d \geq 2$ is the lower bound of $\Omega(nm)$ for $d = 2$ and one-dimensional translations, that we present in Section 5.4.3. Unfortunately this approach does not seem to generalize to higher dimensions, and we are not aware of any other lower bounds.

The only related results in the literature are by Efrat et al. [42] and Venkatasubramanian [72]. In [72] Venkatasubramanian has shown that for the case of polygonal curves in two dimensions and one-dimensional translations along a line Problem 6 can be solved in $O(N^5 \log N)$ time, where N is the sum of the complexities of the two curves. Our approach yields the same runtime. In [42] an algorithm has been presented which solves Problem 7 for the case of two polygonal curves in two dimensions under two-dimensional translations in $O(N^{10}\text{polylog}N)$ time, as opposed to our algorithm which runs in $O(N^8)$ time. However the algorithm of [42] runs in much faster $O(N^4\text{polylog}N)$ time for the variant of Problem 7 when considering the *weak* Fréchet distance. Our algorithm yields only a runtime of $O(N^6 \log N)$ for two-dimensional translations and curves in two dimensions.

Since the algorithms that we present are in general rather time consuming it is interesting to consider approximate algorithms. In [12, 18] it has been shown that for closed convex curves the Hausdorff distance equals the Fréchet distance. This result has been extended to convex surfaces in [45]. In [16] in collaboration with Helmut Alt and Christian Knauer we presented the class of κ-straight curves for which the Fréchet distance is bounded by at most $(\kappa+1)$ times the Hausdorff distance. In [18] we extended this class to the more general class of κ-bounded curves, and showed that we can compute feasible reparameterizations in $O(N \log^2 N 2^{\alpha(N)})$ randomized expected time, where $\alpha(N)$ is the inverse Ackermann function, see [67]. In [17] we have observed, which has independently been observed in [42], that the reference point which maps the two starting points of the curves onto each other is a reference point of quality 1 for translations. This means that mapping the two starting points onto each other yields an approximate algorithm with loss factor 2. This reference point however does not work for closed curves. In [17] we have observed that any reference point for the Hausdorff distance is also a reference point for the Fréchet distance under the same transformation class. Thus any reference point for the Hausdorff distance can be used to approximately match closed curves under the Fréchet distance. See also [56] for most of the results mentioned in this paragraph.

5.1 Basic Properties of the Fréchet Distance

If not stated otherwise, assume that $\varepsilon > 0$ is given and fixed, and let $\rho = ||.||$ be the metric induced by the Euclidean norm. In the sequel we use the notion of a *free space* which was introduced in [10]:

Definition 17 (Free Space / Free Space Diagram) *Let $f : I \to \mathbb{R}^d$, $g : J \to \mathbb{R}^d$ be two curves. The set $F_\varepsilon(f, g) := \{(s, t) \in I \times J \mid ||f(s) - g(t)|| \leq \varepsilon\}$, or F_ε for short, denotes the* free space *of f and g. We call the partition of $I \times J$ into points belonging or not belonging to $F_\varepsilon(f, g)$ the* free space diagram *$FD_\varepsilon(f, g)$ or FD_ε for short.*

We call points in F_ε *white* or *feasible* and points in $FD_\varepsilon \setminus F_\varepsilon$ *black* or *infeasible*. See Figure 5.1 for an illustration. Note that the dimension of the free space of two curves is always two-dimensional, irrespective of the space in which the curves are embedded.

Definition 18 (Cells) *Let $P : [0, m] \to \mathbb{R}^d$, $Q : [0, n] \to \mathbb{R}^d$ be two polygonal curves. We divide the free space diagram $FD_\varepsilon(P, Q)$ into mn* cells *$\zeta_{i,j} := [i, i + 1] \times [j, j + 1]$, for*

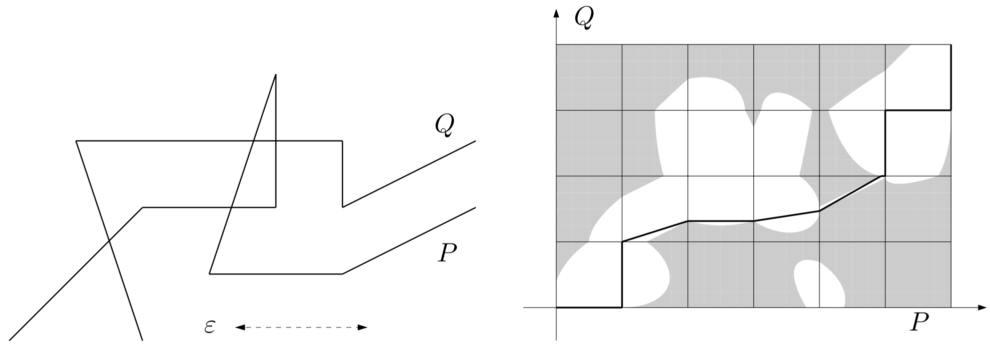

Figure 5.1:
Two polygonal curves P and Q, and their free space diagram. A bi-monotone curve from $(0,0)$ to (m,n) is drawn in the free space. This illustration is taken from [10].

$0 \le i \le m-1$, $0 \le j \le n-1$. *Each vertex* P_i *corresponds to the* vertical line segment $\{i\} \times [0,n]$ *and each* Q_j *to the* horizontal line segment $[0,m] \times \{j\}$ *in the diagram. We call* $[i,i+1] \times [0,n]$ *the* i*-th* column *and* $[0,m] \times [j,j+1]$ *the* j*-th* row.

In Lemma 12 we will see that this grid-like division of the free space diagram is natural for polygonal curves.

The following three lemmata stating basic properties of the free space and its connection to the Fréchet distance are from [10]. Although they are stated there for curves in two dimensions they naturally carry over to curves in higher dimensions. We give the proof of Lemma 11 for d dimensions here, since this lemma needs the most adjustments to carry over to d dimensions. In order to get a better understanding of the topic we also give the proof of Lemma 13. The proof of Lemma 12 is straight-forward, but can also be found in [10].

Lemma 11 *Let* $P, Q : [0,1] \to \mathbb{R}^d$ *be two line segments, and let* $P', Q' : \mathbb{R} \to \mathbb{R}^d$ *be the affine hulls of* P *and* Q*. Then the following holds:*

1. $F_\varepsilon(P', Q')$ *is an elliptical disk (possibly degenerated to the space between two parallel lines) for all* $\varepsilon \ge \rho(P', Q')$*.*
2. $F_\varepsilon(P', Q') = \emptyset$ *for all* $\varepsilon < \rho(P', Q')$ *.*
3. $F_\varepsilon(P, Q)$ *is the intersection of the unit square with* $F_\varepsilon(P', Q')$*.*

Proof: Consider the affine map $f : \mathbb{R}^2 \to \mathbb{R}^d$, $f(s,t) := P'(s) - Q'(t)$. Then, by definition $F_\varepsilon(P', Q') = \{(s,t) \in \mathbb{R}^2 \mid ||f(s,t)|| \le \varepsilon\}$ and $F_\varepsilon(P,Q) = F_\varepsilon(P', Q') \cap [0,1]^2$, which shows the last claim. In order to prove the first two parts, let us first assume that P' and Q' are non-parallel. Then $f(\mathbb{R}^2)$ is a plane in $\mathbb{R}^d$ at distance $\rho(P', Q')$ from the origin. And $I_\varepsilon := \{f(s,t) \mid (s,t) \in \mathbb{R}^2, ||f(s,t)|| \le \varepsilon\}$ is the intersection of $f(\mathbb{R}^2)$ with $\mathbf{B}^d_\varepsilon$, which is empty for $\varepsilon < \rho(P', Q')$, and a disk for $\varepsilon \ge \rho(P', Q')$. Thus $F_\varepsilon(P', Q') = f^{-1}(I_\varepsilon)$ is either empty if $I_\varepsilon = \emptyset$, or it is the the inverse image of a disk under an affine mapping, which is an elliptical disk. If P' and Q' are parallel then $f(\mathbb{R}^2)$ degenerates to a line, such that $F_\varepsilon(P', Q')$ is either empty, or a line. □

For polygonal curves, the pairs of line segments of P and Q naturally subdivide $F_\varepsilon(P, Q)$ into the free spaces for those line segment pairs.

Lemma 12 *Let P, Q be two polygonal curves of complexities m and n, respectively. Then $F_\varepsilon(P,Q)$ is composed of the mn free spaces $F_\varepsilon(\overline{P}_i, \overline{Q}_j) = F_\varepsilon(P,Q) \cap \zeta_{i,j}$ for each pair of line segments $(\overline{P}_i, \overline{Q}_j)$ for $i = 0, \ldots, m-1$ and $j = 0, \ldots, n-1$.*

The reason why we consider the free space diagram in order to compute the Fréchet distance lies in the following lemma.

Lemma 13 *For polygonal curves $P : [0,m] \to \mathbb{R}^d$ and $Q : [0,n] \to \mathbb{R}^d$ we have $\delta_{\mathrm{F}}(f,g) \leq \varepsilon$, iff there exists a curve within $F_\varepsilon(P,Q)$ from $(0,0)$ to (m,n) which is monotone in both coordinates.*

Proof: Let $\gamma : [0,1] \to [0,m] \times [0,n]$, $\gamma(s) = (\gamma_1(s), \gamma_2(s))$, be an (x,y)-monotone curve in $FD_\varepsilon(P,Q)$ from $(0,0)$ to $(1,1)$. γ_1 and γ_2 are monotone reparametrizations of P and Q, respectively. If the whole curve γ is contained in $F_\varepsilon(P,Q)$, then $\delta_{\mathrm{F}}(P,Q) \leq \varepsilon$. In the other direction $\gamma(s) := (\alpha(s), \beta(s))$ defines a monotone curve from $(0,0)$ to (m,n) in $F_\varepsilon(P,Q)$. □

Figure 5.1 shows polygonal curves P, Q, a distance ε, and the corresponding free space diagram FD_ε with the free space F_ε.

Let us now introduce some notation for the different parts of the free space diagram.

Definition 19 (Spikes, Critical Widths and Heights) *Let $(i,j) \in \{0,\ldots,m\} \times \{0,\ldots,n\}$. We denote by*

$L^\varepsilon_{i,j} := \{i\} \times [j + a^\varepsilon_{i,j},\ j + b^\varepsilon_{i,j}]$ *the left line segment, and by*

$B^\varepsilon_{i,j} := [i + c^\varepsilon_{i,j},\ i + d^\varepsilon_{i,j}] \times \{j\}$ *the bottom line segment*

bounding $\zeta_{i,j} \cap F_\varepsilon$. We call the remaining horizontal and vertical parts of the boundary of $\zeta_{i,j}$ that are not contained in F_ε spikes. *In particular*

$\blacktriangle^\varepsilon_{i,j} := \{i\} \times [j, j + a^\varepsilon_{i,j}]$ *is a* lower spike*,*

$\blacktriangledown^\varepsilon_{i,j} := \{i\} \times [j + b^\varepsilon_{i,j}, j+1]$ *an* upper spike*,*

$\blacktriangleright^\varepsilon_{i,j} := [i, i + c^\varepsilon_{i,j}] \times \{j/n\}$ *a* left spike*, and*

$\blacktriangleleft^\varepsilon_{i,j} := [i + d^\varepsilon_{i,j}, i+1] \times \{j/n\}$ *a* right spike.

We call the values $a^\varepsilon_{i,j}$ and $b^\varepsilon_{i,j}$ the critical heights*, and $c^\varepsilon_{i,j}$ and $d^\varepsilon_{i,j}$ the* critical widths *in F_ε. See Figure 5.2 for an illustration.*

Using a dynamic programming approach one can compute those parts of the segments $L^\varepsilon_{i,j}$ and $B^\varepsilon_{i,j}$ which are reachable from $(0,0)$ by a monotone path in F_ε, and thus decide if $\delta_{\mathrm{F}}(P,Q) \leq \varepsilon$ by checking if (m,n) is reachable. See also Section 6.2.1. For details we refer the reader to the proof of the following theorem in [10].

Theorem 7 ([10]) *For given polygonal curves P, Q and $\varepsilon \geq 0$ one can decide in $O(mn)$ time whether $\delta_{\mathrm{F}}(P,Q) \leq \varepsilon$.*

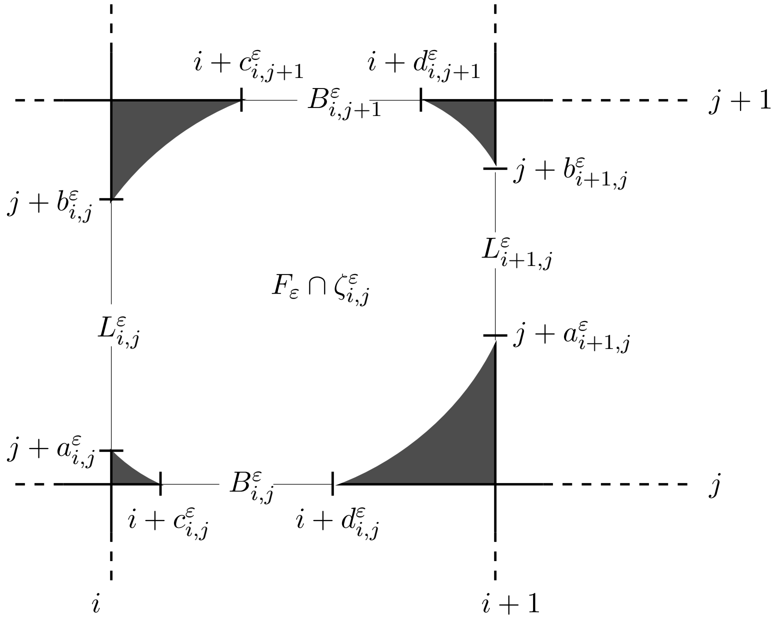

Figure 5.2: Intervals of the free space on the boundary of a cell.

5.2 Polygonal Curves Under Translations

In the following we consider translations $t \in \mathbb{R}^d$ which are applied to the polygonal curve P. For this we adapt the notation from Section 2.1 and from Section 5.1 as follows.

Definition 20 **($P+t$, $F_\varepsilon(t)$, $a^{\varepsilon}_{i,j}(t)$)** *Let $t \in \mathbb{R}^d$ be a translation. Then we denote by $P+t : [0,m] \to \mathbb{R}^d; x \mapsto P(x)+t$ the curve $P : [0,m] \to \mathbb{R}^d$ translated by t, and by $\overline{P}_i(t)$ its ith segment and by $P_i(t)$ its ith vertex. If the context is clear we abbreviate the free space by $F_\varepsilon(t) := F_\varepsilon(P+t, Q)$, and the free space diagram by $FD_\varepsilon(t) := FD_\varepsilon(P+t, Q)$. We denote the cells in $F_\varepsilon(t)$ by $\zeta_{i,j}(t)$, and analogously to Definition 19 the segments bounding a cell by $L^{\varepsilon}_{i,j}(t)$ and $B^{\varepsilon}_{i,j}(t)$, and the critical widths and heights by $a^{\varepsilon}_{i,j}(t)$, $b^{\varepsilon}_{i,j}(t)$, $c^{\varepsilon}_{i,j}(t)$, $d^{\varepsilon}_{i,j}(t)$.*

5.2.1 Continuity

In this subsection we state two insights into the behavior of the free space diagram when it undergoes continuous changes.

Lemma 14 *Let P and Q be two polygonal curves. Then $\delta_{\mathrm{F}}(P+t, Q) \le \delta_{\mathrm{F}}(P,Q) + ||t||$ for all $t \in \mathbb{R}^d$. Moreover, $\delta_{\mathrm{F}}(P+t,Q)$ is continuous as a function in t.*

Proof: For the first claim, consider the reparameterizations α, β that witness $\delta_{\mathrm{F}}(P,Q)$. Clearly α, β are also reparameterizations for $P+t$ and Q, and the claim follows by triangle inequality. The second claim follows directly from the first claim. □

In the following we will consider small changes in ε or t and use the fact that this induces only small changes in $FD_\varepsilon(t)$.

Lemma 15 *Let $t \in \mathbb{R}^d$ be a translation, and $\varepsilon > 0$.*

- *Let $P, Q : \mathbb{R} \to \mathbb{R}^d$ be two lines. Then $F_\varepsilon(P+t, Q)$ varies continuously in ε and t, i.e., $F_\varepsilon(P+t, Q) = \{(s_1, s_2) \in \mathbb{R}^2 \mid g(s_1, s_2, \varepsilon, t) \leq 0\}$ where g is a quadratic polynomial that varies continuously in ε and t.*
- *For two polygonal curves P, Q in $\mathbb{R}^d$ the critical widths and heights $a^\varepsilon_{i,j}(t), b^\varepsilon_{i,j}(t), c^\varepsilon_{i,j}(t), d^\varepsilon_{i,j}(t)$ are continuous functions in ε and t.*

Proof: Let two lines P and Q be given as $P : s_1 \mapsto u_P + s_1 v$ and $Q : s_2 \mapsto u_Q + s_2 w$ with vectors $u_P, u_Q, v, w \in \mathbb{R}^d$ and let $u := u_P - u_Q$. Then $F_\varepsilon(P+t, Q) = \{(s_1, s_2) \in \mathbb{R}^2 \mid ||s_1 v - s_2 w + u + t|| \leq \varepsilon\}$ which can be rewritten as $F_\varepsilon(P+t, Q) = \{(s_1, s_2) \in \mathbb{R}^2 \mid g(s_1, s_2, \varepsilon, t) \leq 0\}$ with $g(s_1, s_2, \varepsilon, t) = v^t v s_1^2 - 2v^t w s_1 s_2 + w^t w s_2^2 + 2s_1 v^t(u+t) - 2s_2 w^t(u+t) + (u+t)^t(u+t) - \varepsilon^2$. This shows that F_ε is a semi-algebraic set defined by the quadratic polynomial $g(s_1, s_2, \varepsilon, t)$, which defines a two-dimensional quadric in (s_1, s_2). Since each quadric class is closed under affine transformations (see e.g. [30]), we conclude from the positive factors in front of s_1^2 and s_2^2 that the quadric is an ellipse for fixed t and ε.

For the second claim let $i \in \{0, 1, ..., m\}$, $j \in \{0, 1, ..., n-1\}$. Let $Q' : s' \mapsto u_Q + s' w$ be the affine hull of $\overline{Q}_j$, and let $P_i = u_P$, and let $u := u_P - u_Q$. Then the roots of the equation $g(0, s', \varepsilon, t) = 0$ with respect to s' are $j + a^\varepsilon_{i,j}(t)$ and $j + b^\varepsilon_{i,j}(t)$ when clipped to $[j, j+1]$. Analogously $i + c^\varepsilon_{i,j}(t)$ and $i + d^\varepsilon_{i,j}(t)$ are also roots of a similar quadratic equation. Thus all critical widths $a^\varepsilon_{i,j}(t), b^\varepsilon_{i,j}(t)$ and heights $c^\varepsilon_{i,j}(t), d^\varepsilon_{i,j}(t)$ are continuous functions in ε and t over the domain for which the roots are real. □

5.2.2 Clampings and Configurations

From Lemma 15 it follows that for two polygonal curves P, Q and translations $t \in \mathbb{R}^d$ the whole free space $F_\varepsilon(P+t, Q)$ varies continuously in ε and t, which we call the *continuity property* of the free space. It implies the existence of the following extreme configurations in the free space diagram. This has been observed in [10] before, however without providing a proof.

Lemma 16 *Let $P : [0, m] \to \mathbb{R}^d$ and $Q : [0, n] \to \mathbb{R}^d$ be two polygonal curves. Let $\varepsilon^* := \delta_\mathrm{F}(P, Q)$. Then for each monotone path π in F_{ε^*} from $(0,0)$ to (m,n) one of the following cases occurs:*

a) There exist indices (i,j) such that either $L^{\varepsilon^*}_{i,j}$ or $B^{\varepsilon^*}_{i,j}$ is a single point on π. (The path passes through a passage between two neighboring cells that consists of a single point.)

b) There exist indices (i,j,k) such that $a^{\varepsilon^*}_{i,j} = b^{\varepsilon^*}_{k,j}$ (or $c^{\varepsilon^*}_{i,j} = d^{\varepsilon^*}_{i,k}$) and π passes through $(i, j + a^{\varepsilon^*}_{i,j})$ and $(k, j + b^{\varepsilon^*}_{k,j})$ (or π passes through $(i + c^{\varepsilon^*}_{i,j}, j)$ and $(i + d^{\varepsilon^*}_{i,k}, k)$)). (The path contains a 'clamped' horizontal or vertical passage, see Figure 5.3.)

c) $L^{\varepsilon^*}_{0,0} = \{(0,0)\}$, $B^{\varepsilon^*}_{0,0} = \{(0,0)\}$, $L^{\varepsilon^*}_{m,n} = \{(m,n)\}$, or $B^{\varepsilon^*}_{m,n} = \{(m,n)\}$.

Proof: By definition there exists a monotone path π from $(0,0)$ to (m,n) in F_{ε^*}. But for all $\varepsilon < \varepsilon^* = \delta_\mathrm{F}(P, Q)$ there does not exist any monotone path from $(0,0)$ to (m,n) in F_ε. There are several cases that can cause this situation: If F_ε is disconnected for all $\varepsilon < \varepsilon^*$, then of course there does not exist *any* path from $(0,0)$ to (m,n) in F_ε. From the continuity of F_ε follows that there must exist indices (i,j) such that $L^\varepsilon_{i,j}$ or $B^\varepsilon_{i,j}$ is a single point, which corresponds to case a). Similarly, if F_ε does not contain $(0,0)$ or (m,n) anymore for all $\varepsilon < \varepsilon^*$ then case c) occurs.

Let us now consider the remaining case, that there does not exist a monotone path from $(0,0)$ to (m,n) in F_ε for all $\varepsilon < \varepsilon^*$ anymore although F_ε is connected and contains $(0,0)$ and (m,n) for $\varepsilon < \varepsilon^*$ close to ε^*. This means that all paths from $(0,0)$ to (m,n) in F_ε must contain a non-monotone part. But since π is monotone in F_{ε^*} and the free space varies continuously in ε we conclude that π must contain a horizontal or vertical piece as stated in case b) which for $\varepsilon < \varepsilon^*$ causes the path to be non-monotone. □

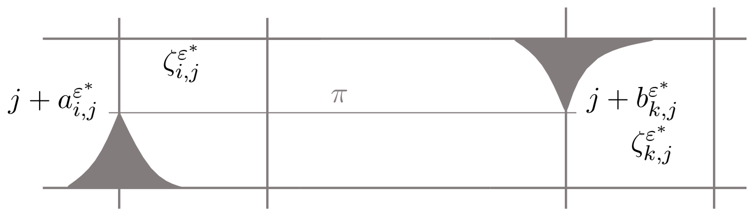

Figure 5.3: The path contains a 'clamped' horizontal passage in the j-th row.

Definition 21 (Clampings) *If a path π in F_ε contains a situation of case a) or b) of Lemma 16 then we say that the path is* clamped *between two spikes. More specifically,*

- *if $a_{i,j}^\varepsilon = b_{k,j}^\varepsilon$, then we call the path is* horizontally clamped *between the lower spike $\blacktriangle_{i,j}^\varepsilon$ and the upper spike $\blacktriangledown_{k,j}^\varepsilon$.*
- *If $c_{i,j}^\varepsilon = d_{i,k}^\varepsilon$ then the path is* vertically clamped *between the left spike $\blacktriangleright_{i,j}^\varepsilon$ and the right spike $\blacktriangleleft_{i,k}^\varepsilon$.*

Case c) is considered as a degenerate subcase of cases a) and b).

Figure 5.4 illustrates the geometric situations that correspond to the clamping situations of Lemma 16. It shows only the horizontal clamping; however the vertical case is symmetric with the roles of P and Q interchanged. In case a) there is only one point $Q(j')$ on $\overline{Q}_j$ at distance ε from P_i, which the reparameterization given by the clamped path in F_{ε^*} maps onto P_i. In case b) there is exactly one point $Q(j')$ on $\overline{Q}_j$ that has distance ε from P_i and P_k. The reparameterization maps $Q(j')$ to every point on the part of P between P_i and P_k. Observe that case a) is a special case of case b) with $i = k$. We include case c) as a special case of both cases a) and b) by considering the endpoints of P and Q as degenerate line segments.

The notion of paths being clamped in the free space diagram is the key concept for the algorithm we present. In the sequel we show that it suffices to consider only those translations that cause the paths in the free space diagram to be clamped at up to d positions. In order to describe possible clamping positions we introduce the notions of configurations and of critical translations:

Definition 22 (Configurations) *A triple (x, y, s) with vertices x, y of one curve and a line segment s of the other curve, is called a* configuration. *In particular, a triple $(P_i, P_k, \overline{Q}_j)$ is called an* h-configuration, *and a triple $(Q_j, Q_k, \overline{P}_i)$ is called a* v-configuration *("h" and "v" stand for horizontal and vertical).*

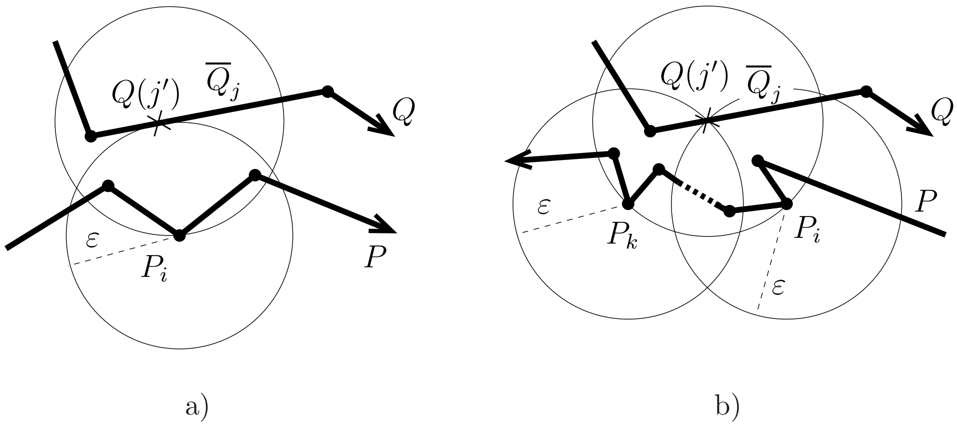

Figure 5.4: The geometric situations corresponding to a horizontally clamped path.

Note that a horizontal clamping of a path which involves indices (i, j, k) corresponds to the h-configuration $(P_i, P_k, \overline{Q}_j)$. Similarly a vertical clamping which involves indices (i, j, k) corresponds to the v-configuration $(Q_j, Q_k, \overline{P}_i)$.

5.2.3 Arrangement of Critical Translations

To a given configuration (x, y, s) we associate the set of translations that give rise TO a potential clamping in that they translate the curve P such that two of the spikes in the free space diagram that correspond to vertices x and y are at the same height in the same row (in case of an h-configuration) or in the same column (in case of a v-configuration).

Definition 23 (Critical Translations) *For an h-configuration $c = (P_i, P_k, \overline{Q}_j)$ with $P_i \neq P_k$ let*

$$T^{\varepsilon}_{crit}(c) := \{t \in \mathbb{R}^d \mid \exists_{z \in \overline{Q}_j} \; ||P_i + t - z|| = \varepsilon \wedge ||P_k + t - z|| = \varepsilon\}$$

be the set of all translations that cause one of the points having distance ε to P_i and P_k to be mapped onto $\overline{Q}_j$.

For $P_i = P_k$ let

$$T^{\varepsilon}_{crit}(c) := \{t \in \mathbb{R}^d \mid \min_{z \in \overline{Q}_j} ||P_i + t - z|| = \varepsilon\}$$

be the set of all translations that cause P_i to have distance ε to $\overline{Q}_j$.

Symmetrically, for a v-configuration $c = (Q_j, Q_k, \overline{P}_i)$ with $Q_j \neq Q_k$ let

$$T^{\varepsilon}_{crit}(c) := \{t \in \mathbb{R}^d \mid \exists_{z \in \overline{P}_i} \; ||Q_j - (z + t)|| = \varepsilon \wedge ||Q_k - (z + t)|| = \varepsilon\}.$$

and for $Q_j \neq Q_k$ let

$$T^{\varepsilon}_{crit}(c) := \{t \in \mathbb{R}^d \mid \min_{z \in \overline{P}_i} ||Q_j - (z + t)|| = \varepsilon\}.$$

We call $T^{\varepsilon}_{crit}(c)$ the set of critical translations *for a configuration c. A translation is called* critical *if it is critical for some configuration.*

The critical translations for a given configuration $c = (x, y, s)$, $x \neq y$, are exactly those translations that preserve the property that the vertices x and y have distance ε to a common point z on the segment s of the other curve. If $x = y$, then the critical translations ensure that there is exactly one point z on the segment s that has distance ε to $x = y$. Note that $x = y$ refers to the position of the vertices x and y, independent of the parameterization within the curve. Thus $x = y$ also covers the case when two different vertices coincide at the same position.

Lemma 17 *Let c be a configuration. Then $T^{\varepsilon}_{crit}(c)$ is a semi-algebraic set of constant description complexity.*

Proof: Let $c = (x, y, s)$, $x \neq y$. The definition of $T^{\varepsilon}_{crit}(c)$ involves a first-order boolean formula, i.e., a boolean formula with one existential quantifier. We square the equations in order to obtain polynomials without square roots. From the technique of quantifier elimination, see [31], we know that there exists an equivalent boolean formula without an existential quantifier. In the case that $x = y$ the definition of $T^{\varepsilon}_{crit}(c)$ consists, after squaring the equation, of an expression of the form $\min_z p(t, z) = \varepsilon$ for a polynomial p in t and z. Clearly we can rewrite this expression as $\exists_{z^*} \forall_z : p(t, z^*) = \varepsilon \wedge p(t, z) \geq p(t, z^*)$. This is again a first-order boolean formula for which by quantifier elimination there exists an equivalent boolean formula without any quantifiers. Thus in both cases $T^{\varepsilon}_{crit}(c)$ is indeed a semi-algebraic set. It is of constant description complexity since the number of involved polynomials is always constant, and the quantifier elimination keeps this number constant, too. □

Now that we know that $T^{\varepsilon}_{crit}(c)$ is a semi-algebraic set, let us consider which surface it describes in $\mathbb{R}^d$. By reformulating Definition 23 we obtain the following characterizations:

Observation 2 *For an h-configuration $c = (P_i, P_k, \overline{Q}_j)$ with $P_i \neq P_k$*

$$\begin{aligned} T^{\varepsilon}_{crit}(c) &= \{t \in \mathbb{R}^d \mid \exists_{z \in \overline{Q}_j}\, t - z \in -(S^{d-1}_{\varepsilon}(P_i) \cap S^{d-1}_{\varepsilon}(P_k))\} \\ &= \overline{Q}_j \oplus (-(S^{d-1}_{\varepsilon}(P_i) \cap S^{d-1}_{\varepsilon}(P_k)))\ , \end{aligned}$$

and for $P_i = P_k$

$$T^{\varepsilon}_{crit}(c) = \partial(\overline{Q}_j \oplus (-\mathbf{B}^d_{\varepsilon}(P_i)))\ .$$

For a v-configuration $c = (Q_j, Q_k, \overline{P}_i)$ with $Q_j \neq Q_k$

$$\begin{aligned} T^{\varepsilon}_{crit}(c) &= \{t \in \mathbb{R}^d \mid \exists_{z \in \overline{P}_i}\, t + z \in S^{d-1}_{\varepsilon}(Q_j) \cap S^{d-1}_{\varepsilon}(Q_k))\} \\ &= -\overline{P}_i \oplus ((S^{d-1}_{\varepsilon}(Q_j) \cap S^{d-1}_{\varepsilon}(Q_k)))\ , \end{aligned}$$

and for $Q_j = Q_k$

$$T^{\varepsilon}_{crit}(c) = \partial(-\overline{P}_i \oplus (-\mathbf{B}^d_{\varepsilon}(Q_j)))\ .$$

In two dimensions for an h-configuration $c = (P_i, P_k, \overline{Q}_j)$ for example, $T^{\varepsilon}_{crit}(c)$ consists of two parallel copies of $\overline{Q}_j$ in distance $2\sqrt{\varepsilon^2 - ||P_i - P_k||^2/4}$ if $P_i \neq P_k$. If $P_i = P_k$ then those two segments are joined at both ends by a half-circle of radius ε. For $d = 3$, $T^{\varepsilon}_{crit}(c)$ is a finite portion of a cylinder which, if $P_i = P_k$, is closed by two half-spheres of radius ε at both ends.

Let $c = (P_i, P_k, \overline{Q}_j)$ be an h-configuration with $P_i = P_k$. We can describe $T^\varepsilon_{crit}(c)$ by the following polynomials: Let $p_{\text{cyl}}(t)$ be the quadratic polynomial in t such that $p_{\text{cyl}}(t) = 0$ describes the hollow infinite cylinder of radius ε around the affine hull of $\overline{Q}_j - P_i$. Let $h_1(t)$ and $h_2(t)$ be two hyperplanes normal to $\overline{Q}_j - P_i$, such that $(Q_j - P_i) \in h_1(t)$ and $(Q_{j+1} - P_i) \in h_2(t)$, and both halfspaces $h_1(t) \leq 0$ and $h_2(t) \leq 0$ contain $\overline{Q}_j - P_i$. Let $p_{\text{sphere1}}(t) = 0$ be the sphere of radius ε with center $Q_j - P_i$, and let $p_{\text{sphere2}}(t) = 0$ be the sphere of radius ε with center $Q_{j+1} - P_i$. Then

$$\begin{aligned} T^\varepsilon_{crit}(c) \;\; = \{t \in \mathbb{R}^d \mid \;\; & (\,(p_{\text{cyl}}(t) = 0) \wedge (h_1(t) \leq 0) \wedge (h_2(t) \leq 0)\,) \\ & \vee (\,(p_{\text{sphere}\,1}(t) = 0) \wedge (h_1(t) \geq 0)\,) \\ & \vee (\,(p_{\text{sphere}\,2}(t) = 0) \wedge (h_2(t) \geq 0)\,)\,\} \end{aligned}$$

If $P_i \neq P_k$ then we know from Observation 2 that

$$\begin{aligned} T^\varepsilon_{crit}(c) &= \overline{Q}_j \oplus (-(S^{d-1}_\varepsilon(P_i) \cap S^{d-1}_\varepsilon(P_k))) \\ &= \{t \in \mathbb{R}^d \,|(\,(p_{\text{cyl}}(t) = 0) \wedge (h_1(t) \leq 0) \wedge (h_2(t) \leq 0)\,)\,\}\,, \end{aligned}$$

where $p_{\text{cyl}}(t)$ describes a hollow infinite cylinder around the affine hull of $\overline{Q}_j - P_i$. However its radius is not ε but it depends on the angle between the normal hyperplane to $\overline{Q}_j$ and the normal hyperplane to $P_i - P_k$. $h_1(t)$ and $h_2(t)$ describe two cut-off hyperplanes normal to $P_i - P_k$ such that $(Q_j - (P_i + P_k)/2) \in h_1(t)$ and $(Q_{j+1} - (P_i + P_k)/2) \in h_2(t)$.

Now let us turn to a key property of critical translations.

Lemma 18 *Let $c = (P_i, P_k, \overline{Q}_j)$ be an h-configuration. Then $a^\varepsilon_{i,j}(t) \leq b^\varepsilon_{k,j}(t)$ for all $t \in T^\varepsilon_{crit}(c)$.*

Let $c = (Q_j, Q_k, \overline{P}_i)$ be a v-configuration. Then $c^\varepsilon_{j,i}(t) \leq d^\varepsilon_{k,i}(t)$ for all $t \in T^\varepsilon_{crit}(c)$.

Proof: Let $c = (P_i, P_k, \overline{Q}_j)$ be an h-configuration and assume that there exists a $t' \in T^\varepsilon_{crit}(c)$ with $a^\varepsilon_{i,j}(t') > b^\varepsilon_{k,j}(t')$. From the definition of the free space we know that $a^\varepsilon_{i,j}(t) \leq b^\varepsilon_{i,j}(t)$ as well as $a^\varepsilon_{k,j}(t) \leq b^\varepsilon_{k,j}(t)$, for all $t \in \mathbb{R}^d$. Thus we have $a^\varepsilon_{k,j}(t') \leq b^\varepsilon_{k,j}(t') < a^\varepsilon_{i,j}(t') \leq b^\varepsilon_{i,j}(t')$. On the other hand, for all critical translations $t \in T^\varepsilon_{crit}(c)$ there exist by definition an $h \in \{a^\varepsilon_{i,j}, b^\varepsilon_{i,j}\}$ and an $h' \in \{a^\varepsilon_{k,j}, b^\varepsilon_{k,j}\}$ such that $h(t) = h'(t)$. This is however not possible for t'. Contradiction. The argument for a v-configuration is analogous. □

In other words Lemma 18 states that the spikes associated with a configuration c define (when taking no other spikes into account) a passage in the free space diagram that for all $t \in T^\varepsilon_{crit}(c)$ never forces a possible path to change its orientation from monotone increasing to monotone decreasing.

Let us now consider the arrangement in translation space that is naturally formed by all critical translations.

Definition 24 *The set of all $T^\varepsilon_{crit}(c)$ for all possible configurations c form an arrangement A^ε_{crit} in translation space $\mathbb{R}^d$. We call A^ε_{crit} the* arrangement of critical translations.

From Observation 2 and Lemma 17 it follows that A^ε_{crit} is an arrangement of patches of algebraic surfaces of constant description complexity involving polynomials of degree at most two. In two dimensions, for example, A^ε_{crit} is consists of line segments and circular arcs.

In the sequel we will see that we only have to consider translations that lie on any of the surfaces forming the arrangement, but not inside its cells.

Lemma 19 *Let $t \in \mathbb{R}^d$ be a translation. If $\delta_{\mathrm{F}}(P+t,Q)=\varepsilon$, then t is critical.*

Proof: By Lemma 16 there is a path π in $F_\varepsilon(P+t,Q)$ for which one of the cases a), b), or c) occurs. If the corresponding geometric situation (see Figure 5.4) involves the vertices $P_i(t)$ and $P_k(t)$ on $P+t$ and a point on the edge $\overline{Q}_j(t)$ then t is critical for the h-configuration $(P_i, P_k, \overline{Q}_j)$. If the geometric situation involves vertices from Q and a segment of $P+t$ the same argument yields a v-configuration. □

Note that the condition in Lemma 19 is clearly only sufficient and not necessary.

Lemma 20 *If there are translations $t_<, t_> \in \mathbb{R}^d$ such that $\delta_{\mathrm{F}}(P+t_<,Q) < \varepsilon < \delta_{\mathrm{F}}(P+t_>,Q)$ then there is a critical translation $t_= \in \mathbb{R}^d$ such that $\delta_{\mathrm{F}}(P+t_=,Q)=\varepsilon$.*

Proof: By continuity of F_ε, see Lemma 15, there exists a transformation $t_=$ on any curve between $t_<$ and $t_>$ in translation space such that $\delta_{\mathrm{F}}(P+t_=,Q)=\varepsilon$. By Lemma 19 the translation $t_=$ is critical. □

Since there is always a translation which brings P within Fréchet distance greater than ε to Q, we now know from Lemma 20 that, in order to search for a translation $t \in \mathbb{R}^d$ with $\delta_{\mathrm{F}}(P+t,Q) \le \varepsilon$, it suffices to restrict our search to all critical translations that bring P in Fréchet distance exactly ε to Q. However, from Observation 2 we know that the critical translations form $(d-1)$-dimensional surfaces in $\mathbb{R}^d$, so there remains an infinitely large set of possible translations. But we will see that it suffices to consider only one translation per face of the arrangement of critical translations, of which there are only finitely many.

Lemma 21 *If there is a translation $t_< \in \mathbb{R}^d$ such that $\delta_{\mathrm{F}}(P+t_<,Q) < \varepsilon$ then there is some k-dimensional face F in A^ε_{crit}, $0 \le k \le d-1$, such that $\delta_{\mathrm{F}}(P+t,Q) \le \varepsilon$ for all $t \in F$.*

Proof: Clearly there exists a translation t_* such that $\delta_{\mathrm{F}}(P+t_*,Q) > \varepsilon$. By Lemma 20 there is a configuration c and a critical translation $t_= \in T^\varepsilon_{crit}(c)$ such that $\delta_{\mathrm{F}}(P+t_=,Q)=\varepsilon$. Let F be the connected component of $T^\varepsilon_{crit}(c)$ such that $t_= \in F$. If $\delta_{\mathrm{F}}(P+t,Q) \le \varepsilon$ for all $t \in F$ we are done. Otherwise there exists a $t_> \in F$ with $\delta_{\mathrm{F}}(P+t_>,Q) > \varepsilon$.

Now consider an arbitrary curve τ on F such that $\tau(0)=t_=$ and $\tau(1)=t_>$. The existence of such a curve is proven for example in [31] (Proposition 2.5.13). By continuity there has to be an $s^* \in [0,1]$ such that $\delta_{\mathrm{F}}(P+\tau(s^*),Q)=\varepsilon$ and $\delta_{\mathrm{F}}(P+\tau(s),Q) > \varepsilon$ for all $s > s^*$ arbitrary close to s^*. For clarity of exposition assume that $s^*=0$. Considering the free space diagram, this means that $F_\varepsilon(P+t_=,Q)$ contains a monotone path, but $F_\varepsilon(P+\tau(s),Q)$ does not contain this or any other monotone path anymore for all $s>0$ that are arbitrary close to 0.

By continuity of F_ε this can only happen if there are two critical heights $a^\varepsilon_{i,j}, b^\varepsilon_{i',j}$ (or critical widths $c^\varepsilon_{i,j}, d^\varepsilon_{i,j'}$) such that

- $a^\varepsilon_{i,j}(t_=) = b^\varepsilon_{i',j}(t_=)$ (or $c^\varepsilon_{i,j}(t_=) = d^\varepsilon_{i,j'}(t_=)$),
- each monotone path in $F_\varepsilon(P+t_=,Q)$ passes horizontally between $(i, j+a^\varepsilon_{i,j}(t_=))$ and $(i', j+b^\varepsilon_{i',j}(t_=))$ (or vertically between $(i+c^\varepsilon_{i,j}(t_=), j)$ and $(i+d^\varepsilon_{i,j'}(t_=), j')$),
- and $a^\varepsilon_{i,j}(\tau(s)) > b^\varepsilon_{i',j}(\tau(s))$ (or $c^\varepsilon_{i,j}(\tau(s)) > d^\varepsilon_{i,j'}(\tau(s))$) for $s>0$ arbitrary close to 0.

In other words all monotone paths in $F_\varepsilon(P+t_=,Q)$ pass horizontally between the lower spike $\blacktriangle^\varepsilon_{i,j}$ and the upper spike $\blacktriangledown^\varepsilon_{i',j}$ (or vertically between $\blacktriangleright^\varepsilon_{i,j}$ and $\blacktriangleleft^\varepsilon_{i,j'}$). And for $s>0$ the spikes close the only possible monotone passage.

Let the h-configuration $c' = (P_i, P_{i'}, \overline{Q}_j)$ (or v-configuration $c' = (Q_j, Q_{j'}, \overline{P}_i)$) be the configuration that corresponds to the two spikes causing the clamping.

We know from Lemma 18 that, in case of an h-configuration $a^\varepsilon_{i,j}(\tau(s)) \le b^\varepsilon_{k,j}(\tau(s))$, and in case of a v-configuration $c^\varepsilon_{j,i}(\tau(s)) \le b^\varepsilon_{k,i}(\tau(s))$ for all $s \in [0,1]$, since $\tau(s) \in T^\varepsilon_{crit}(c)$. Thus the critical heights / widths corresponding to C have the same height / width for all $\tau(s)$, i.e., they cannot close the passage. Thus $c \neq c'$, and $t_= \in T^\varepsilon_{crit}(c) \cap T^\varepsilon_{crit}(c')$.

Now consider the $(d-2)$-dimensional face $f \subseteq T^\varepsilon_{crit}(c) \cap T^\varepsilon_{crit}(c')$ of A^ε_{crit} which contains $t_=$. We apply the same argumentat as above inductively. This either yields a vertex t of A^ε_{crit} with $\delta_{\mathrm{F}}(P+t,Q)=\varepsilon$, or, if the process ends early, a k-dimensional face f of A^ε_{crit}, $1 \le k \le d-1$, such that $\delta_{\mathrm{F}}(P+t,Q) \le \varepsilon$ for all $t \in f$. Hence, the claim follows by induction.

□

Note that we do not have to assume general position for the proof of Lemma 21. In fact, it is possible that two (or more) k-faces of A^ε_{crit} intersect in a k-dimensional manifold. However if this happens then we know from Lemma 18 that neither of the spikes corresponding to those k-faces can close the passage in the free space diagram, such that there has to be another set of critical translations intersecting the k-faces, and thus reducing the dimension by one.

By Lemma 21 we know that it suffices to construct a point on each face of A^ε_{crit}, of which there are finitely many, and to check the Fréchet distance for exactly these translations.

Theorem 8 *Let P, Q be two polygonal curves in $\mathbb{R}^d$ of complexities m and n, respectively, let $N := m+n$, and let $\varepsilon > 0$. Then one can solve Problem 7 for the set of translations, i.e., one can find a translation $t \in \mathbb{R}^d$ such that $\delta_{\mathrm{F}}(P+t,Q) \le \varepsilon$, if it exists, in $O((nm)^{d+1}(n+m)^d) = O(N^{3d+2})$ time and $O((nm(n+m))^d) = O(N^{3d})$ space.*

Proof: According to Lemma 21 it suffices to consider one point in each face of the arrangement A^ε_{crit} of critical translations. In the notation of [25, 27] these points form what is called a *semi-algebraic sample*. Note that the number of different configurations, and thus the number of algebraic surfaces forming the arrangement A^ε_{crit} is $M := O(nm(n+m))$. We use a result of Basu, Pollack, and Roy [25, 27] which states that such a semi-algebraic sample has size $O(M^d)$ and can be computed in $O(M^{d+1})$ time and $O(M^d)$ space. However we can avoid a linear factor in M for the computation as follows: Basu, Pollack, and Roy consider all k-tuples of the M polynomials, for $1 \le k \le d$.[†] Each such k-tuple defines a semi-algebraic set. For each such k-tuple they compute a set of points such that each face of the semi-algebraic set defined by the k-tuple is intersected by at least one of the points. This is called the *Sample Points Subroutine*. For this they first compute a *k-univariate representation* of the desired set of sample points, from which they afterwards compute the actual points. A k-univariate representation of a set of points S is a $(k+2)$-tuple $u := (f, g_0, \ldots, g_k)$, where $f, g_0, \ldots, g_k$ are univarirate polynomials in $x \in \mathbb{R}$, such that $S \subseteq \mathrm{Assoc}(u)$ where $\mathrm{Assoc}(u) := \{(g_1(\alpha)/g_0(\alpha), \ldots, g_k(\alpha)/g_0(\alpha)) \mid \alpha \text{ is a real root of } f\}$ is the set of *associated points* of u. For one k-tuple they generate a constant number of k-univariate representations

[†]In fact for general position arguments they consider for each polynomial four perturbed versions of it, and choose from the set of the $4M$ perturbed polynomials. For the sake of clarity of presentation we do not consider this larger set of polynomials, although our arguments carry directly over to this set.

in constant time [‡]. Thus the computation of all k-univariate representations for all k-tuples of polynomials, $1 \leq k \leq d$ takes $O(M^d)$ time in total. The step to compute the points from a k-univariate representation (which is called the *Univariate Sign Determination Subroutine* in [27]) adds an overhead of $O(M)$ per k-univariate representation, thus introducing a multiplicative factor M to the total. This subroutine however computes all roots of f (and thus also all the desired sample points), and for each (real) root the sign of every of the M input polynomials at the point associated with this root. But for our algorithm we do not need the sign information of the sample points. The pure computation of the sample points without the sign information for every polynomial takes only constant time. Thus the whole computation of the sample points can be done in $O(1)$ time per k-univariate representation, summing to $O(M^d)$ in total.

Having computed the $O(M^d)$ sample points we check for each of these critical translations in $O(mn)$ time, see Theorem 7 and [10], whether it yields a distance $\leq \varepsilon$. Observe that space for the last step has to be allocated only once. Thus the runtime is $O(M^d mn) = O((nm)^{d+1}(n+m))^d)$ and the space complexity is $O(M^d)$. □

5.2.4 Parametric Search

In order to find a translation that minimizes the Fréchet distance between the two polygonal curves we apply the parametric search paradigm. We make use of several related papers by Basu, Pollack, and Roy on arrangements of semi-algebraic surfaces [26, 28, 24, 27].

Consider all polynomial equations and inequalities that define all critical translations $T^{\varepsilon}_{crit}(c)$ for all configurations c. We denote with $\mathcal{P}_{=}$ the set of all polynomials that define such an equation, with $\mathcal{P}_{\leq}$ the set of polynomials that define an inequality, and with $\mathcal{P} := \mathcal{P}_{=} \cup \mathcal{P}_{\leq}$ their union. We assume without loss of generality that the inequalities are $P' \leq 0$ for all $P' \in \mathcal{P}_{\leq}$. All polynomials $P' \in \mathcal{P}$ are polynomials $P'(\varepsilon, t)$ in $t \in \mathbb{R}^d$ and in $\varepsilon \in \mathbb{R}$. They are all quadratic in ε, with no linear terms in ε, and with no mixed terms in t and ε. The polynomials in $\mathcal{P}_{=}$ are quadratic polynomials in $t \in \mathbb{R}^d$ that describe spheres or surfaces of cylinders. Polynomials in $\mathcal{P}_{\leq}$ are affine in $t \in \mathbb{R}^d$ and describe hyperplanes that bound half-spaces. For a fixed $\varepsilon' > 0$ and a polynomial $P' \in< \mathcal{P}$ we denote by $P'(\varepsilon') := P'(\varepsilon', t)$ the polynomial in $t \in \mathbb{R}^d$ for fixed $\varepsilon = \varepsilon'$. Similarly we denote by $\mathcal{P}(\varepsilon')$ the set of all polynomials in $\mathcal{P}$ for fixed $\varepsilon = \varepsilon'$. [§]

Let again $M := |\mathcal{P}| = O(nm(n+m))$. Essentially we construct points on every face of A^{ε}_{crit} as we did in the proof of Theorem 8. Then we "track" those points for varying ε, maintaining those $O(M)$ sample points as functions in ε. For each such function (which corresponds to a translation depending on ε) we consider the critical values, and perform a similar parametric search as in [10].

Note that A^{ε}_{crit} is the semi-algebraic set defined by the polynomials in $\mathcal{P}$.

We apply similar results as in the proof of Theorem 8, but parameterized by the "special" parameter ε. First of all, let us define a *parameterized k-univariate representation*. This is a $(k+2)$-tuple $u := (f, g_0, \ldots, g_k)$, where $f, g_0, \ldots, g_k$ are polynomials in $t \in \mathbb{R}$ pand $\varepsilon \in \mathbb{R}$. For each $\varepsilon' \in \mathbb{R}$, the set of *associated points* is the set $\mathrm{Assoc}^{\varepsilon'}(u) :=$

[‡]Note that the degrees of the polynomials are bounded by d, and as everywhere in this thesis d is assumed to be constant.

[§]Note that we identify $P'(\varepsilon, t_1, \ldots, t_d)$ with $P'(\varepsilon, (t_1, \ldots, t_d)) = P'(\varepsilon', t)$ for $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$ and $\varepsilon \in \mathbb{R}$.

$\{(g_1(\alpha,\varepsilon)/g_0(\alpha,\varepsilon),\ldots,g_k(\alpha,\varepsilon)/g_0(\alpha,\varepsilon)) \mid \alpha$ is a real root of $f\}$. We apply the *Parameterized Sample Points Subroutine*, see [27], on $\mathcal{P}_= \cup \mathcal{P}_\leq$ which computes in $O(M^d)$ time a set $\mathcal{U}$ of k-univariate representations such that for every $\varepsilon' \in \mathbb{R}$ the union of points $\bigcup_{u\in\mathcal{U}} \mathrm{Assoc}^{\varepsilon'}(u)$ intersects every face of $\mathcal{P}(\varepsilon')$.¶ For each k-univariate representation $u \in \mathcal{U}$ we apply the *Parameterized Real Roots Subroutine* of [28] (see also [24]). For a fixed $u = (f, g_0, \ldots, g_k) \in \mathcal{U}$ it computes in constant time a partition of the ε-axis $\mathbb{R}$ into intervals such that on each interval the number of real roots of f is the same. For each interval endpoint ε' of this partition all real roots $\alpha(\varepsilon')$ of $f(\varepsilon')$, which correspond to points in different connected components on $A^{\varepsilon'}_{crit}$, are computed. Likewise, for every interval in the partition a semi-algebraic curve segment $\sigma(\varepsilon) := (\frac{g_1(\varepsilon,\alpha(\varepsilon))}{g_0(\varepsilon,\alpha(\varepsilon))},\ldots,\frac{g_k(\varepsilon,\alpha(\varepsilon))}{g_0(\varepsilon,\alpha(\varepsilon))})$ is computed, for ε varying in the interval. This curve segment continuously "tracks" a sample point for varying ε, until the sample point changes the connected component of $A^{\varepsilon'}_{crit}$. We determine adjacencies between points and curve segments as is done in [28]. This way we obtain a graph in $\mathbb{R}^{d+1}$ with semi-algebraic arcs as edges. Altogether the time to compute these $O(M^d)$ graphs is $O(M^d)$. This is a simpler variant of the *road map* for semi-algebraic surfaces, which is computed in [28, 24]. The road map involves more complex structures since it forces the sub-road map in every connected component in $\mathbb{R}^{d+1}$ to be connected. The advantage of our case is that we can consider the different parameterized k-univariate presentations independently and do not have to worry about crossings or connectivity issues.

With the same argument as in the proof of Lemma 21 we know that if $\varepsilon^* := \min_{t\in\mathbb{R}^d} \delta_{\mathrm{F}}(P+t,Q)$, then there is some k-dimensional face F in $A^{\varepsilon^*}_{crit}$, $0 \leq k \leq d-1$, such that $\delta_{\mathrm{F}}(P+t,Q) = \varepsilon^*$ for all $t \in F$. Thus, in order to find the optimal ε^* we have to check points on every face of $A^{\varepsilon^*}_{crit}$. But how do we find ε^*? We consider the $O(M^d)$ graphs. For each such graph we consider each of its semi-algebraic curve segments, of which there are only $O(1)$ many. Each such curve segment corresponds to a critical translation $t(\varepsilon)$ which is parameterized by ε on one of the intervals on the partition of the ε-axis that is computed in the *Parameterized Real Roots Subroutine* of [28, 24]. We apply the same parametric search as in [10] on $P+t(\varepsilon)$ and Q. The only difference is that $P+t(\varepsilon)$ is not a fixed polygonal curve but varies depending on ε. For the parametric search however the critical widths and heights are considered depending on ε anyhow. So the dependency on ε for the polygonal curve in the first place does not affect the general concept. The parametric search in [10] for two fixed polygonal curves of complexity m and n can be done in $O(mn\log(mn))$ time. This approach utilizes Cole's trick for parametric search based on sorting which yields a logarithmic instead of poly-logarithmic overhead in the runtime. Thus, since we consider $O(M^d)$ curve segments and we apply the parametric search on the translation and the range of ε which is defined by each segment, we obtain a $O(M^d mn\log(mn))$ time algorithm to actually compute the minimum Fréchet distance under translations, and thus to solve Problem 6 for translations. This proves the following theorem:

Theorem 9 *Let P, Q be two polygonal curves in $\mathbb{R}^d$ of complexities m and n, respectively, let $N := m+n$. Then we can solve Problem 6 for translations, i.e., we can find $\varepsilon^* > 0$ and a translation $t^* \in \mathbb{R}^d$ such that $\varepsilon^* = \delta_{\mathrm{F}}(P+t^*,Q) = \min_{t\in\mathbb{R}^d}\delta_{\mathrm{F}}(P+t,Q)$ in $O((nm)^{d+1}(n+m)^d\log(mn)) = O(N^{3d+2}\log N)$ time and $O((nm(n+m))^d) = O(N^{3d})$ space.*

¶Note that in [27] the polynomials in the computed parameterized k-univariate representations contain infinitely large and small variables. However these can be reduced at asymptotically no extra cost as it is done in the *Parameterized Cell Representatives Subroutine* by applying the *Parameterized Removal of Infinitely Large Variables*.

5.3 Polygonal Curves Under Affine Transformations

The approach of Section 5.2 can be generalized to more general motions than translations. A rather general form is to assume that we are given a subset $\mathcal{T}$ of the affine transformations in $\mathbb{R}^d$.[‖] Each affine transformation is composed of a linear transformation and a translation, and can thus be described by a pair (A, t) where A is a $d \times d$ matrix representing the linear transformation and t is a d-dimensional translation. Thus a general affine transformation in $\mathbb{R}^d$ has $d^2 + d$ parameters. Note that A could be the identity matrix, or t could be the zero vector.

5.3.1 Arrangement of Critical Transformation Parameters

We need to restrict ourselves to subsets of the affine transformations that are represented in a way such that we can apply the approach from Section 5.2.

Definition 25 *Let $1 \leq k \leq d^2 + d$, and let $p_1, \ldots, p_{d^2+d}, q_1, \ldots, q_{d^2+d} \in \mathbb{R}[X_1, \ldots, X_k]$ be $2(d^2 + d)$ polynomials of constant degree in k variables*[**]*, such that $q_i(x) \neq 0$ for all $1 \leq i \leq d^2 + d$ and for all $x \in \mathbb{R}^k$. Let $r_i := p_i/q_i$ for all $1 \leq i \leq d^2 + d$, such that $r_i(x) := p_i(x)/q_i(x)$ for all $x \in \mathbb{R}^k$. If*

$$\mathcal{T} = \left\{ \left(\begin{pmatrix} r_1(x) & \ldots & r_d(x) \\ r_{d+1}(x) & \ldots & r_{2d}(x) \\ \vdots & & \vdots \\ r_{d^2-d}(x) & \ldots & r_{d^2}(x) \end{pmatrix}, \begin{pmatrix} r_{d^2+1}(x) \\ r_{d^2+2}(x) \\ \vdots \\ r_{d^2+d}(x) \end{pmatrix} \right) \middle| \; x \in \mathbb{R}^k \right\}$$

then we call $\mathcal{T}$ rationally parameterized *or* rationally represented *with k* degrees of freedom (dof)*. $\mathbb{R}^k$ is called the* parameter space *of $\mathcal{T}$.*

In the rest of this section we assume that $\mathcal{T}$ is rationally parameterized. This model is general enough in that it covers the most commonly used transformation classes, such as lower dimensional rotations, rigid motions, and similarities, and arbitrary affine transformations for example.

We generalize the techniques of Section 5.2 to any rationally parameterized subset $\mathcal{T}$ of the affine transformations. We will see that the runtime of our algorithm depends on the number of degrees of freedom of the representation. All definitions and most of the lemmata carry over to this more general setting. However this time we identify transformations in $\mathcal{T}$ with their parameters in $\mathbb{R}^k$. More specifically, for any $x \in \mathbb{R}^k$ we denote by $(A_x, t_x) \in \mathcal{T}$ its corresponding transformation in $\mathcal{T}$. Similar to Definition 23 we define the set of *critical transformation parameters* for every configuration.

Definition 26 (Critical Transformation Parameters) *Let $\mathbb{R}^k$ be the parameter space of $\mathcal{T}$. For an h-configuration $c = (P_i, P_k, \overline{Q}_j)$ with $P_i \neq P_k$ let*

$$T^{\varepsilon}_{crit}(c) := \{x \in \mathbb{R}^k \mid \exists_{z \in \overline{Q}_j} \; ||A_x P_i + t_x - z|| = \varepsilon \wedge ||A_x P_k + t_x - z|| = \varepsilon\}$$

be the set of parameters of all transformations that cause one of the points having distance ε to P_i and P_k to be mapped onto $\overline{Q}_j$.

[‖] In principle our constructions also carry over to projective transformations.

[**] $\mathbb{R}[X_1, \ldots, X_k]$ is the ring of all polynomials in k variables with real coefficients.

For $P_i = P_k$ let

$$T^{\varepsilon}_{crit}(c) := \{x \in \mathbb{R}^k \mid \min_{z \in \overline{Q}_j} ||A_x P_i + t_x - z|| = \varepsilon\}$$

be the set of parameters of all transformations that cause P_i to have distance ε to $\overline{Q}_j$.

Symmetrically, for a v-configuration $c = (Q_j, Q_k, \overline{P}_i)$ with $Q_j \neq Q_k$ let

$$T^{\varepsilon}_{crit}(c) := \{x \in \mathbb{R}^k \mid \exists_{z \in \overline{P}_i} ||Q_j - (A_x z + t_x)|| = \varepsilon \wedge ||Q_k - (A_x z + t_x)|| = \varepsilon\}.$$

and for $Q_j \neq Q_k$ let

$$T^{\varepsilon}_{crit}(c) := \{x \in \mathbb{R}^k \mid \min_{z \in \overline{P}_i} ||Q_j - (A_x z + t_x)|| = \varepsilon\}.$$

We call $T^{\varepsilon}_{crit}(c)$ the set of critical transformation parameters *for a configuration c. A transformation parameter is called* critical *if it is critical for some configuration.*

Note that in Section 5.2 for the case of translations we implicitly also worked on the parameter space, since we assumed the set of all translations to be represented in a way such that the parameter space equaled the translation space.

In the following we carry over the results of Section 5.2 step by step:

Lemma 22 *For each configuration c the set $T^{\varepsilon}_{crit}(c)$ is a semi-algebraic set of constant description complexity in $\mathbb{R}^k$.*

Proof: Let first c be a configuration with two different vertices. The definition of $T^{\varepsilon}_{crit}(c)$ involves one existential quantifier. The vectors inside the norm consist of weighted sums of polynomials r_i for $1 \leq i \leq d^2 + d$, where the weights or coefficients depend on c. In fact we can equally consider the square of the norm instead of the norm, and when we multiply by all involved q_i, $1 \leq i \leq d^2+d$, we obtain a boolean formula involving polynomials in x of constant degree (assuming that d is constant). This boolean formula is preceded by one existential quantifier, and thus represents a first-order boolean formula. Using quantifier elimination [27] we can compute in constant time an equivalent quantifier-free boolean formula involving a constant number of polynomials in x of constant degree.

A similar argument holds for the case when the two vertices in c are equal. Considering the square of the norm and again multiplying by all involved q_i, $1 \leq i \leq d^2 + d$, we obtain an expression of the form $\min_z p(x, z) = \varepsilon$ for a polynomial p in x and z. The same argument as in the proof of Lemma 17 shows that this is expression is equivalent to a boolean formula involving a constant number of polynomials in x of constant degree. Thus in both cases $T^{\varepsilon}_{crit}(c)$ is a semi-algebraic set of constant description complexity. □

For each $x \in \mathbb{R}^k$ we denote by $F_{\varepsilon}(x) := F_{\varepsilon}(A_x P + t_x, Q)$, and the free space diagram by $FD_{\varepsilon}(x) := FD_{\varepsilon}(A_x P + t_x, Q)$. We denote the cells in $F_{\varepsilon}(x)$ by $\zeta_{i,j}(x)$, and analogously to Definition 19 the segments bounding a cell by $L^{\varepsilon}_{i,j}(x)$ and $B^{\varepsilon}_{i,j}(x)$, and the critical widths and heights by $a^{\varepsilon}_{i,j}(x)$, $b^{\varepsilon}_{i,j}(x)$, $c^{\varepsilon}_{i,j}(x)$, $d^{\varepsilon}_{i,j}(x)$.

The next lemma is analogous to Lemma 18, and the proof carries directly over, such that we omit it.

Lemma 23 *Let $c = (P_i, P_k, \overline{Q}_j)$ be an h-configuration. Then $a^{\varepsilon}_{i,j}(x) \leq b^{\varepsilon}_{k,j}(x)$ for all $x \in T^{\varepsilon}_{crit}(c)$.*

Let $c = (Q_j, Q_k, \overline{P}_i)$ be a v-configuration. Then $c^{\varepsilon}_{j,i}(x) \leq b^{\varepsilon}_{k,i}(x)$ for all $x \in T^{\varepsilon}_{crit}(c)$.

Definition 27 *The set of all $T^{\varepsilon}_{crit}(c)$ for all possible configurations c form an arrangement A^{ε}_{crit} in $\mathbb{R}^k$. We call A^{ε}_{crit} the* arrangement of critical transformation parameters.

If there does not exist an $x_> \in \mathbb{R}^k$ such that $\delta_{\mathrm{F}}(A_{x_>}P + t_{x_>}, Q) > \varepsilon$, then trivially $\delta_{\mathrm{F}}(A_x P + t_x, Q) \leq \varepsilon$ for any $x \in A^{\varepsilon}_{crit}$. On the other hand, if there exists an $x_> \in \mathbb{R}^k$ such that $\delta_{\mathrm{F}}(A_{x_>}P + t_{x_>}, Q) > \varepsilon$, then Lemma 19 and Lemma 20 carry directly over to our more general setting. Also Lemma 21 carries over in the same way. The proof uses properties of semi-algebraic sets, and the continuity of the free space under the transformations. Taking a look at the proof of Lemma 15 we see that a similar argument, involving the same reasoning as in the proof of Lemma 22, also yields the continuity in our more general setting. Thus we know that the following holds:

Lemma 24 *If there is a transformation $x_< \in \mathbb{R}^k$ such that $\delta_{\mathrm{F}}(A_{x_<}P + t_{x_<}, Q) < \varepsilon$ then there is some k'-dimensional face F in A^{ε}_{crit}, $0 \leq k' \leq k-1$, such that $\delta_{\mathrm{F}}(A_x P + t_x, Q) \leq \varepsilon$ for all $x \in F$.*

In fact, taking a look at the proofs of Theorem 8 and Theorem 9 we see that all arguments involve arrangements of semi-algebraic sets, and no special properties of the translation space. Thus both proofs again carry directly over, of course working in k-dimensional parameter space $\mathbb{R}^k$ now instead of the d-dimensional translation space $\mathbb{R}^d$. Thus we obtain the following theorem:

Theorem 10 *Let P, Q be two polygonal curves in $\mathbb{R}^d$ of complexities m and n, respectively, let $N := m + n$, Furthermore, let $\mathcal{T}$ be a subset of affine transformations that is rationally parameterized by the parameter space $\mathbb{R}^k$ with k degrees of freedom.*

For given $\varepsilon > 0$ one can solve Problem 7, i.e., one can find a transformation $(A, t) \in \mathcal{T}$ such that $\delta_{\mathrm{F}}(AP+t, Q) \leq \varepsilon$, in $O((nm)^{k+1}(n+m)^k) = O(N^{3k+2})$ time and $O((nm(n+m))^k) = O(N^{3k})$ space.

Using parametric search one can solve Problem 6, i.e., one can find $\varepsilon^* > 0$ and a transformation $x^* \in \mathbb{R}^k$ such that $\varepsilon^* = \delta_{\mathrm{F}}(A_{x^*}P + t_{x^*}, Q) = \min_{x \in \mathbb{R}^k} \delta_{\mathrm{F}}(A_x P + t_x, Q)$ in $O((nm)^{k+1}(n+m)^k \log(mn)) = O(N^{3k+2} \log N)$ time and $O((nm(n+m))^k) = O(N^{3k})$ space.

Note that k can be as large as $d^2 + d$, which of course depends on the choice of the rationally parameterized $\mathcal{T}$ of the affine transformations.

5.3.2 Specific Transformation Classes

Now let us give example rational parameterizations for commonly used sets of transformations.

Rotations

Let first $d = 2$. Any rotation matrix in two dimensions has the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

with $a^2 + b^2 = 1$, see [29] for example. The standard parameterization is to set $a = \cos\alpha$ and $b = \sin\alpha$ for a radian angle $\alpha \in [0, 2\pi)$. This parameterization however does not fulfill our requirements of a rational parameterization. Therefore we use the following different parameterization

$$a(x) := \frac{2x}{1+x^2} \qquad b(x) := \frac{1-x^2}{1+x^2}$$

for $x \in \mathbb{R}$, see for example [60]. We denote by A_x^{rot} the corresponding rotation matrix, and by $\mathcal{T}_{\mathrm{rot}}^2 := \{A_x^{\mathrm{rot}} \mid x \in \mathbb{R}\}$ the set of all rotation matrices generated by all possible parameters. Note that this parameterization describes the rotations with angles in $(-\pi/2, 3\pi/2)$. The angle $3\pi/2$ is excluded, which is however not a real restriction, since any angle arbitrarily close to $3\pi/2$ is included. We consider in the following $\mathcal{T}_{\mathrm{rot}}^2$ as the set of all possible two-dimensional rotations. This parameterization fulfills the requirements of Definition 25, and is thus a rational parameterization of $\mathcal{T}_{\mathrm{rot}}^2$ with 1 degree of freedom.

One standard approach for $d = 3$ is to write each three-dimensional rotation matrix as the product of three rotation matrices, where each of the three matrices corresponds to a rotation by some angle around one of the three axes. Those angles are called the Euler angles, see [60] for example. Each of these three matrices corresponds essentially to a two-dimensional rotation matrix. Applying the rational parameterization for each of these matrices, and then multiplying the matrices together we obtain a rational parameterization of the space of three-dimensional rotations. This rational parameterization has 3 degrees of freedom. We denote the set of three-dimensional rotations generated this way $\mathcal{T}_{\mathrm{rot}}^3$.

In general, a rotation in d dimensions is specified by an orthonormal matrix. There are $d + d(d-1)/2$ conditions that specify the orthonormality of a $d \times d$ matrix, such that $d(d-1)/2$ parameters remain free to choose, see [29] for example. Unfortunately we are not aware of a rational parameterization for $d \geq 4$, which we could plug into Theorem 10. In fact be aware that our parameterizations for $\mathcal{T}_{\mathrm{rot}}^2$ and $\mathcal{T}_{\mathrm{rot}}^3$ only parameterize almost all rotations. See [60] for a more comprehensive study of different parameterizations for rotations and rigid motions. However the rational parameterization given there that fully covers $\mathcal{T}_{\mathrm{rot}}^3$ has four instead of three degrees of freedom. It would be interesting to be aware of higher-dimensional rational parameterizations, or to use a different approach to represent affine transformations for our application.

Rigid Motions

Rigid motions are a combination of rotations with translations. Since the translations have a natural rational parameterization with d degrees of freedom, we know that every rigid motion in d dimensions has $d(d-1)/2 + d = d(d+1)/2$ degrees of freedom. Thus we know that for two-dimensional rigid motions there is a rational parameterization with 3 degrees of freedom, and for the three-dimensional rigid motions with 6 degrees of freedom.

Scalings

A scaling is the multiplication of the vector by a fixed scalar value $\lambda \in \mathbb{R}$. Thus the affine transformation is λ times the identity matrix, which is clearly a rational parameterization with 1 degree of freedom. Thus there is a rational parameterization for d-dimensional scalings with 1 degree of freedom.

Similarities

Similarities are combinations of rigid motions with scalings. Thus a d-dimensional similarity has $d(d+1)/2 + 1 = d(d+3)/2$ degrees of freedom. Thus for two-dimensional similarities there is a rational parameterization with 4 degrees of freedom, and for the three-dimensional rigid motions with 7 degrees of freedom.

Affine Transformations

Arbitrary affine transformations in d dimensions have a trivial rational parameterization with $d^2 + d$ degrees of freedom, by specifying all $d^2 + d$ entries in the matrix and the translation vector.

5.4 Variants

5.4.1 Weak Fréchet Distance

Consider the matching problems Problem 6 and Problem 7 to be defined for the *weak* Fréchet distance, see Definition 9. In this case we can apply almost the same techniques as in Section 5.2 and Section 5.3. Consider Lemma 16 in this case. For the weak Fréchet distance we drop the monotonicity condition of a path in the free space from $(0,0)$ to (m,n). Thus for two polygonal curves P and Q in $\mathbb{R}^d$, and $\varepsilon^* := \tilde{\delta}_{\mathrm{F}}(P,Q)$, a clamping of a path π in F_{ε^*} can only occur in cases *a*) and *c*) of Lemma 16. These types of clampings are defined by a segment of one curve and a vertex of the other curve. For the configurations this means that we only consider *weak configurations* which are either $(P_i, \overline{Q}_j)$ for a vertex P_i of P and a segment $\overline{Q}_j$ of Q, or $(Q_j, \overline{P}_i)$ for a vertex Q_i of Q and a segment $\overline{P}_j$ of P. In fact, those weak configurations are a subset of the regular configurations. Notice that in Definition 23 and Definition 26 we handled the special cases where the two vertices of the configuration are the same. These special cases correspond exactly to the sets of *weak critical translations* or *weak critical transformation parameters* for the weak configurations. Hence, the weak critical transformation parameters are a subset of the critical transformation parameters. The remaining argument is exactly the same as in Section 5.2 and Section 5.3. But since we are now only considering weak configurations, the arrangement of critical transformation parameters consists only of $O(mn)$ surfaces, instead of $O(mn(n+m))$ many. Thus the number of necessary sample points in the arrangement of critical transformation parameters is $O((mn)^k)$, such that we obtain the following variant of Theorem 10:

Corollary 8 *Let P, Q be two polygonal curves in $\mathbb{R}^d$ of complexities m and n, respectively, let $N := m + n$, Furthermore, let $\mathcal{T}$ be a subset of affine transformations that is rationally parameterized by the parameter space $\mathbb{R}^k$ with k degrees of freedom.*

For given $\varepsilon > 0$ one can solve Problem 7 for the weak Fréchet distance, i.e., one can find a transformation $(A,t) \in \mathcal{T}$ such that $\tilde{\delta}_{\mathrm{F}}(AP + t, Q) \leq \varepsilon$, in $O((nm)^{k+1}) = O(N^{2k+2})$ time and $O(nm) = O(N^{2k})$ space.

Using parametric search one can solve Problem 6 for the weak Fréchet distance, i.e., one can find $\varepsilon^* > 0$ and a transformation $x^* \in \mathbb{R}^k$ such that $\varepsilon^* = \tilde{\delta}_{\mathrm{F}}(A_{x^*}P + t_{x^*}, Q) = \min_{x\in\mathbb{R}^k} \tilde{\delta}_{\mathrm{F}}(A_x P + t_x, Q)$ in $O((nm)^{k+1}\log(mn)) = O(N^{2k+2}\log N)$ time and $O(nm) = O(N^{2k})$ space.

Note that for curves in two dimensions and two-dimensional translations the result of [42] outperforms our result by roughly a quadratic factor. It would be interesting to see if the results of [42] could be generalized to work in our more general setting.

5.4.2 Closed Curves and Partial Matching

In this subsection we consider two generalizations that have already been considered in [10]: Closed curves and partial curve matching.

Closed Curves

So far we have considered only *open curves*, i.e., curves defined as mappings over an interval. If we alter Definition 4 to consider continuous mappings defined over the circle $S^1_{\mathbf{1}}$, then we arrive at the notion of *closed curves*. In other words we can define a closed curve f as a curve $f : [0,1] \to \mathbb{R}^d$ with $f(0) = f(1)$. For a closed curve we do not want to treat 0 and 1 as special start or end points. For the Fréchet distance we would rather allow the "man" and the "dog" to pick the best possible starting points. For $a \in [0,1]$ we define the *shift of f by a* as $f_a : [a, 1+a] \to \mathbb{R}^d$ with $f_a(x) = f(x \mod 1)$. Then the extension of the Fréchet distance to closed curves has been defined in [10] as follows:

Definition 28 *Let $f, g : [0,1] \to \mathbb{R}^d$ be two closed curves. Then*

$$\delta_{\mathrm{F}}(f,g) := \min_{a,b\in[0,1]} \delta_{\mathrm{F}}(f_a, g_b)$$

It has been shown in [10] that the Fréchet distance can be computed for closed polygonal curves P and Q of complexities m and n, respectively, in $O(mn\log^2(mn))$ time. The decision problem, whether the Fréchet distance is at most ε can be solved in $O(mn\log(mn))$ time. In order to accomplish this, Alt et al.[10] construct a reachability data structure on top of the free space diagram for P and Q. This data structure allows to test whether there exists a monotone path in the free space from a given point on the lower boundary to another given point on the upper boundary of the free space diagram. Then they utilize this data structure to check all possible different shifts of P. The clamping that changes the existence of monotone paths in the free space are the same as in Lemma 16[††]. This means that our approach can directly be applied to closed curves to compute a finite set of critial transformation parameters. For each of these transformations we now apply the decision algorithm of [10] for closed curves. The parametric search also carries over to this setting. We thus obtain the same runtimes as for the case of open curves, however with an additional $O(\log(mn))$ factor.[‡‡]

Corollary 9 *Let P, Q be two closed polygonal curves in $\mathbb{R}^d$ of complexities m and n, respectively, let $N := m + n$, Furthermore, let $\mathcal{T}$ be a subset of affine transformations that is rationally parameterized by the parameter space $\mathbb{R}^k$ with k degrees of freedom.*

For given $\varepsilon > 0$ one can solve Problem 7 for closed curves, i.e., one can find a transformation $(A,t) \in \mathcal{T}$ such that $\tilde{\delta}_{\mathrm{F}}(AP + t, Q) \le \varepsilon$, in $O((mn)^{k+1}\log(mn)) = O(N^{2k+2}\log N)$ time and $O((mn) = O(N^{2k}\log N)$ space.

[††]Note that in [10] there are new types of clampings introduced. However the configuration corresponding to such a clamping defines a set of critical translation parameters which equals the set of critical translation parameters for one of the configurations that we already consider. Thus we do not need to include these new clampings.

[‡‡]This factor stems from the divide and conquer construction of the reachability data structure in [10].

The corresponding optimization problem for closed curves Problem 6 can be solved in $O((mn)^{k+1}\log^2(mn)) = O(N^{2k+2}\log^2 N)$ time and $O((mn) = O(N^{2k})$ space.

Partial Matching

Another interesting variant of matching two curves is the partial matching task: Given two polygonal curves P and Q, find a subcurve P' of P such that $\delta_F(P', Q)$ is minimized. A subcurve is a restriction of the supercurve to a closed interval. This task has been attacked in [10] by utilizing the same reachability data structure as for the case of closed curves, and hence obtaining the same runtimes as for the case of closed curves.

For this we can just as for the case of closed curves apply our techniques in the same way to find a transformation $(A, t) \in \mathcal{T}$ and a subcurve P' of P such that $\delta_F(AP' + t, Q)$ is minimized. Again we first compute the finite set of critial transformation parameters, and then apply the $O(mn\log(mn))$ decision algorithm of [10]. The parametric search also carries over again.

Corollary 10 *Let P, Q be two polygonal curves in $\mathbb{R}^d$ of complexities m and n, respectively, let $N := m + n$. Furthermore, let $\mathcal{T}$ be a subset of affine transformations that is rationally parameterized by the parameter space $\mathbb{R}^k$ with k degrees of freedom.*

For given $\varepsilon > 0$ one can decide if there exists a transformation $(A, t) \in \mathcal{T}$ and a subcurve P' of P such that $\delta_F(AP' + t, Q) \leq \varepsilon$ in $O((mn)^{k+1}\log(mn)) = O(N^{2k+2}\log N)$ time and $O((mn) = O(N^{2k})$ space.

Using parametric search one can solve the corresponding minimization problem in $O((nm)^{k+1}\log^2(mn)) = O(N^{2k+2}\log^2 N)$ time and $O((nm) = O(N^{2k})$ space.

5.4.3 A Lower Bound

The algorithms we presented in this chapter have a very high time complexity, which raises the question if it is really necessary to construct the arrangement of so many different critical transformation parameters. We would like to know how complex the arrangement of critical transformation parameters can actually be – it might be that our upper bounds are way too high. Unfortunately, we can only present a lower bound for curves in $\mathbb{R}^2$ and one-dimensional translations, which does not seem to easily generalize to higher dimensions.

Let $\varepsilon > 0$ be given. We construct two polygonal curves P and Q in $\mathbb{R}^2$ of complexities m and n, respectively, such that for horizontal translations $t \in \mathbb{R}$ the distance between $P + t$ and Q alternates $\Omega(nm)$ times between being less than ε and larger than ε. Thus the combinatorial complexity of A^{ε}_{crit} for one-dimensional translations (thus $k = 1$) for curves in $d = 2$ dimensions is $\Omega(nm)$. Let $b := \varepsilon/n$, and choose $0 < w < a$ such that $a + w < b/m$. We construct P and Q to consist of m and n vertical spikes of height ε, width w, and side-length $d := \sqrt{\varepsilon^2 + w^2/4}$. In particular, P starts with a horizontal line segment of length ε, followed by m downward spikes which are connected by short horizontal line segments of length a, and followed by another horizontal line segment of length $\varepsilon - b$. Q starts with a horizontal line segment of length b, followed by n upward spikes spaced in distance b which are connected by short horizontal line segments of length b. See Figure 5.5. We let the horizontal parts of P and Q be positioned on top of each other. Then for any horizontal translation t, if every downward spike of P is completely below a horizontal segment of Q and every upward spike of Q is completely above a horizontal segment of P then $\tilde{\delta}_F(P + t, Q) = \delta_F(P + t, Q) = \varepsilon$. If however for a translation t' a downward spike is positioned right below an upward spike, then $\tilde{\delta}_{\mathrm{F}}(P+t',Q) = \delta_{\mathrm{F}}(P+t',Q) = d > \varepsilon$.

Notice that this is only a lower bound for the complexity of the arrangement of critical transformation parameters. A completely different approach to solve Problem 7 does not have to be related to this complexity.

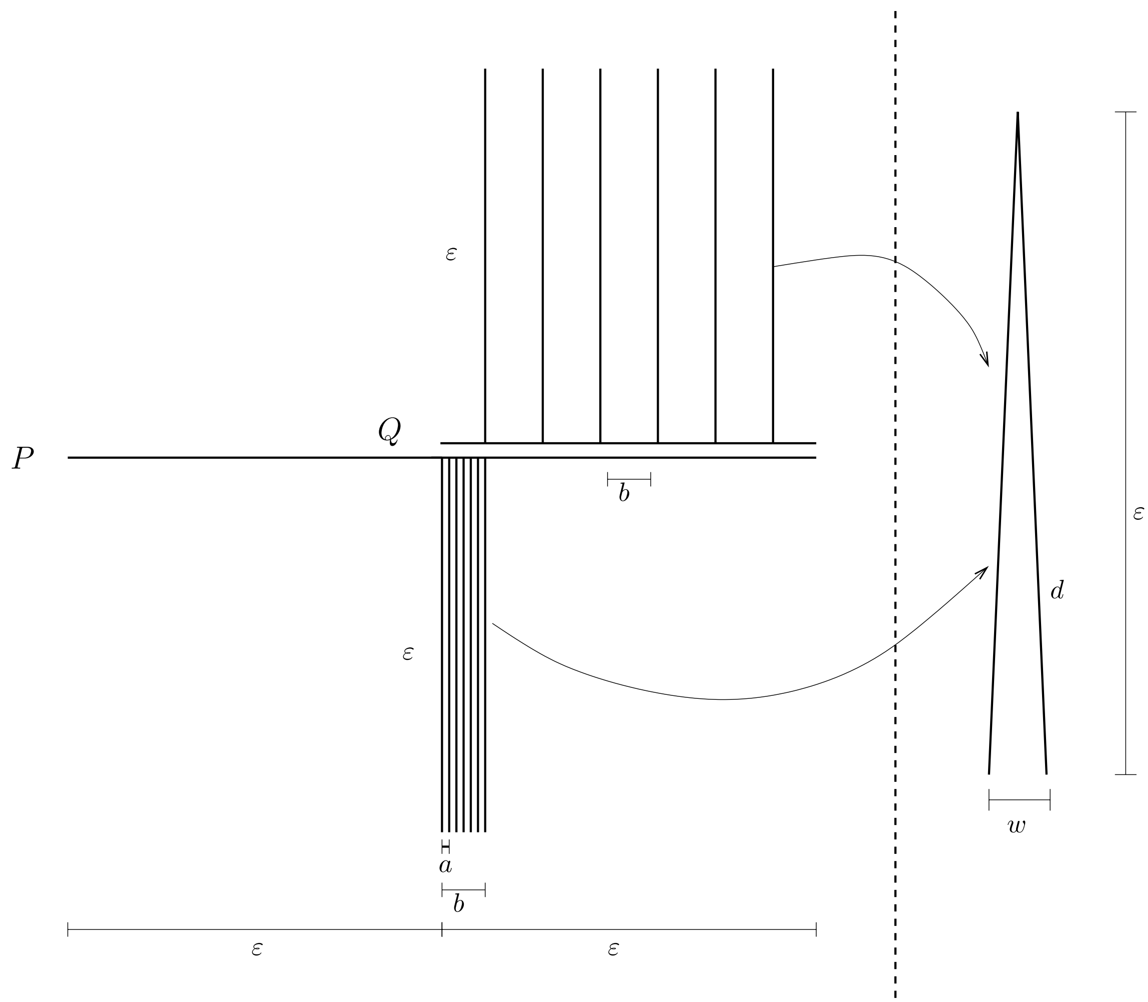

Figure 5.5: The $\Omega(nm)$ lower bound for the Fréchet distance under translations.