How to take short cuts

Claire Kenyon
LIENS URA 1327, Ecole Normale Supérieure,
45 rue d'Ulm, 75230 Paris Cedex 05, France.

Richard Kenyon
IHES, 35 route de Chartres, 91440 Bures-sur-Yvette, France.

1 Introduction

Given a smooth curve or polygon C of length 1 in the plane, can one add "short cuts" to C of small total length so that any two points on C are within distance (along the curve and short cuts) at most a constant factor away from their Euclidean distance?

Peter Jones has solved this problem in the affirmative [1]:

Theorem 1 (P. Jones) *There are constants C_0, C_1 such that for every polygon P there is a rectifiable set Γ, $P \subset \Gamma$ with total length $L(\Gamma) \leq C_0 L(P)$ satisfying: if $x, y \in P$ there is a subarc $\gamma \subset \Gamma$ from x to y with total length $L(\gamma) \leq C_1|x-y|$.*

Actually he proves this (and more) for P any connected rectifiable set. His proof, however, uses some difficult techniques from harmonic analysis, and the construction in [1] is not necessarily computable.

We supply here a geometric proof of the above theorem, along with a simple construction and algorithm using the *skeleton* of the polygon P, a subgraph of the Voronoi diagram of the edges.

Our construction necessarily adds an infinite number of segments, but in practice when one only needs finite precision, it is finite. The construction also generalizes to smooth curves and planar graphs. It can be implemented easily, the two major subroutines being a convex hull algorithm and a Voronoi diagram algorithm (for edges).

In section 2, we describe the contruction of the short cuts across the polygon. In section 3, we give an algorithm for finding a short path (using P and the short cuts) between any pair of points on the boundary of P. In section 4, we analyze the total length of the short cuts added, and show that it is at most a constant times the length of P. In section 5, given a pair of points on the boundary, we analyze the length of the path found by our algorithm, and show that it is at most a constant times their Euclidian distance. Most of the proofs of the lemmas use elementary geometry only: they are only sketched here and will be detailed in the full paper.

2 The Construction.

We need to construct a family of segments linking together points of the polygon which are too far apart. We first construct the convex hull of P. This divides the plane into several connected regions. All the finite regions (which we call R_i) are polygons.

We construct short cuts in each region separately. We first construct the **skeleton** of the polygonal boundary of R_i, defined as the set of centers of circles contained in R_i and which touch the boundary at two or more points. Such circles are called **Voronoi circles**.

It is easy to see that the skeleton is a tree whose leaves are exactly the convex vertices of the boundary, and whose branching points have degree 3 exactly except in degenerate cases (the skeleton is a subgraph of the Voronoi diagram of the edges of the boundary). Its edges are formed with pieces of straight lines and pieces of parabolas: straight lines for the centers of all the Voronoi circles which pass through two concave vertices of the boundary, and for the centers of all the Voronoi circles which are tangent to two edges of the boundary; parabolas for the centers of all the Voronoi circles which go through a concave vertex and are tangent to an edge.

Let a be a constant greater than 1. Our first operation is called "trimming the acute leaves". For each convex vertex v of the boundary such that the angle at v is smaller than $2Arc\ sin(1/a)$, we consider the straight edge of the skeleton leading to that vertex, and choose some point c_0 on that edge. It is the center of a circle tangent at points p_0 and q_0 to the two edges adjacent to v. We add segment p_0q_0 to the list of short cuts. We move from c_0 on the skeleton towards v, adding segments p_iq_i to the list of short cuts, defined inductively

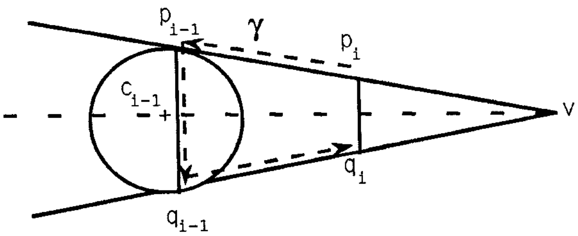

Figure 1: Trimming the acute vertices

as follows: given $c_{i-1}, p_{i-1}, q_{i-1}$, we move from c_{i-1} on the skeleton towards v. Any point c corresponds to two points p and q on the boundary; let $\gamma(p,q)$ be the length of the path

$$p \to p_{i-1} \to q_{i-1} \to q$$

going from p to p_{i-1} along the boundary, along the short cut $p_{i-1}q_{i-1}$, and from there along the boundary to q. Let $|p-q|$ be the Euclidian distance between p and q. As we move towards v, $\gamma(p,q)$ increases and $|p-q|$ decreases. We define c_i to be the point such that $\gamma(p_i,q_i) = a|p_i - q_i|$. See figure 1 for an illustration. Thus we add an infinite sequence of cuts which get shorter and shorter as we get closer to v. We can now erase the part of the skeleton from c_0 to v. After this is done for every acute vertex of the skeleton, we are finished with trimming.

We choose to root the trimmed skeleton at an arbitrary point on an edge leading to a leaf. The main part of our construction of short cuts will be in bottom-up order from the leaves to the root of the trimmed skeleton, erasing the skeleton as we move up it. In other words, we only deal with a branching point once all its descendants have been dealt with (and so the skeleton below the branching point is all erased), and we deal with an edge once its endpoint furthest from the root has been erased by walking along the edge from that endpoint towards the root, erasing the edge as we move along.

Let p and q be two points of the original polygon. A **quick path** (not to be confused with **short path**, which we define later) $\gamma(p,q)$ between p and q is a path along the boundary and the short cuts, with total length at most $a|p-q|$.

As we are doing our construction, we keep the following invariant property: if c, a point on the skeleton, has been erased, then there are quick paths between the boundary points on the Voronoi circle of center c. We observe that this property is indeed true right after the original trimming: the convex vertices with small angle have been dealt with precisely so that the property would be satisfied.

We now need to define how we deal with an edge or with a branching point of skeleton. If c is our current position, two cases arise: either the circle associated to c touches the boundary at exactly two points p and q ,

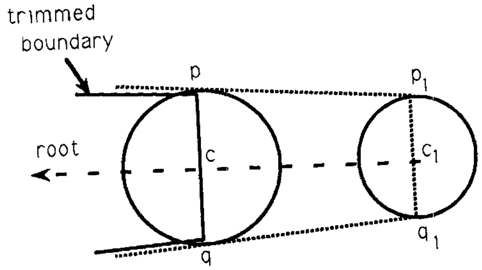

Figure 2: Constructing the short cuts: case 1

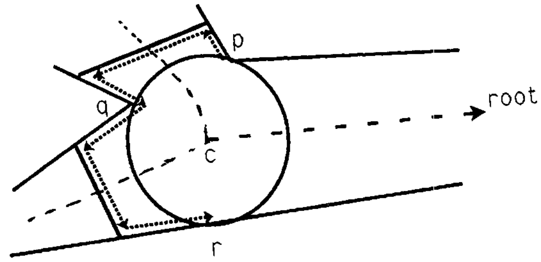

Figure 3: Constructing the short cuts: case 2

or c is a vertex of the skeleton and touches the boundary at three points p, q and r.

Case 1: dealing with an edge (c_1, c_2) in the skeleton. Let c_1 be the endpoint furthest from the root. c_1 is already erased, so there are quick paths between the boundary points p_1, q_1, r_1 on its Voronoi circle. As we move from c_1 to c_2, let p, q be our current boundary points, and $\gamma(p,q)$ the path from p to q which goes through p_1 and q_1. If $\gamma(p,q)$ is a quick path, we erase c and keep going. If $\gamma(p,q)$ equals $a|p-q|$, we add segment pq to the list of short cuts, and trim the boundary further by replacing the part followed by $\gamma(p,q)$ by the segment pq. See figure 2.

Case 2: dealing with a branching point: The Voronoi circle at c touches the boundary at p, q, r. Suppose that the root of the skeleton is "between" p and r. We know that there is already a quick path between p and q and between q and r. Let $\gamma(p,r)$ be the concatenation of those two paths. As in the first case, if $\gamma(p,r) \geq a|p-r|$ we add segment pr to the list of short cuts and trim the boundary by replacing $\gamma(p,r)$ by segment pr; otherwise we do nothing. Then we erase the branching point. See figure 3.

For the construction to be well defined, we only need to show that the process in the main part of the construction is finite.

Lemma 1 *The process is finite.*

Proof: If we remove small neighborhoods around each vertex, any two points on the polygon are either on the same edge or at distance at least ϵ apart, $\epsilon > 0$. Thus the segments we add during the main part of the construction all have length at least ϵ. Every time we add a segment, the boundary length is trimmed by at least $(a-1)\epsilon$. Since it has finite length, we can only add a finite number of segments. □

We cannot bound the number of segments added independently of the polygon; in fact no such bound holds for *any* short cut construction: a polygon which has two parallel edges very close to each other must have a large number of short cuts. If we are operating under the assumption of finite resolution, however, we ignore edges of length less than ϵ, and this lemma gives a bound on the number of other edges needed since the total length is bounded.

After this construction is done, all the skeleton is erased, and so, for any two points p and q which are on the same Voronoi circle, there is a quick path between p and q. In the next section, we will forget about the actual construction of the short cuts and only use that property: whenever we are on a Voronoi circle, we can hop across to any other point on the circle with cost at most equal to a times the Euclidian distance spanned.

3 Finding the short path.

Once we have constructed the short cuts of P, we need to describe how to get quickly from one point on P to another.

Let l be the line segment between two points p and q of the polygon. We can assume that l does not cross the boundary of P, since in that case consider the short paths between successive intersections of P with l; the concatenation of these short paths gives a short path from p to q.

Let $c(p)$ be the center of the Voronoi circle $C(p)$ which touches p (in the region R containing l). Similarly define $c(q)$. Let $V(p,q)$ denote the unique shortest path in the skeleton between $c(p)$ and $c(q)$. Since p and q see each other, the following property holds:

Lemma 2 *$V(p,q)$ is monotone increasing from $c(p)$ to $c(q)$, in direction $q-p$.*

For any path γ from p to q, contained in region R, there is a corresponding path γ_R from p to q along the boundary and short cuts of R; we define γ_R in the following way.

The skeleton partitions R into (closed) regions which we call **skeleton cells**. We have the initial points $\gamma_R(0) = \gamma(0) = p$. Let x_0 be the first point at which γ leaves the skeleton cell $S(p)$ containing p; let C_0 be the Voronoi circle with center x_0. Then C_0 contains at least two points of the boundary of R, and exactly one point x_p in the cell $S(p)$. Define the initial segment of γ_R to be the unique path on R which goes from p to that point x_p, and is contained in the region $S(p)$.

The path γ moves from x_0 into a new skeleton cell S', and C_0 has exactly one point x' on R in this cell.

By the construction of our short cuts, there is a quick path from x_p to x'; extend γ_R by this path. This defines γ_R from p to x', and by repeating the same construction from cell to cell we define γ_R for the entire path γ.

To define a **short path** from p to q, we consider first the segment l from p to q. Unfortunately, the corresponding path γ_l around R constructed in the above manner may be very inefficient (see figure 7). However the inefficiencies are only caused by l intersecting $V(p,q)$ a large number of times at small angle; we will perturb l slightly, creating a new path $\tilde{l}$ for which $\gamma_{\tilde{l}}$ is more efficient.

Let α be a given angle, a constant of our construction. We define $\tilde{l}$ so that it follows l as closely as possible, but is constrained to lie on one side of $V(p,q)$, unless $V(p,q)$ is too steep.

More precisely, $\tilde{l}$ starts at p, and agrees with l until the first intersection point, x, of $V(p,q)$ with l. We can assume (by perturbing p or q slightly) that x is not a vertex of V; since the $V(p,q)$ at x is then smooth, we can measure the angle of intersection of $\tilde{l}$ and $V(p,q)$ at x. If this angle is larger than α, our predefined constant, we let $\tilde{l}$ cross $V(p,q)$ and continue in direction l into the new region.

On the other hand, if this angle is smaller than α, we let $\tilde{l}$ follow $V(p,q)$, diverging from l, so as to remain on the same side of $V(p,q)$. We let $\tilde{l}$ continue to follow $V(p,q)$ (remaining just slightly to one side) until either we arrive at $c(q)$, arrive at l again, or $V(p,q)$ starts diverging from l with angle larger than α.

In the first case, we complete $\tilde{l}$ by simply going straight to q. Since $c(q)$ is the center of $C(q)$, this path lies within R, and does not encounter $V(p,q)$ again.

In the second case, we let $\tilde{l}$ follow l further and react similarly at later intersections.

In the third case, we let $\tilde{l}$ cross $V(p,q)$ at that point where the angle becomes greater than α, and then let it go *straight towards* q, until it reaches q, or again encounters $V(p,q)$. In the latter case we follow $V(p,q)$ without crossing it until we reach l again or else $V(p,q)$ points upward away from q, (in which case we continue going straight towards q).

This completely defines the path $\tilde{l}$ from p to q, and (by the construction after Lemma 2) a path $\gamma_{\tilde{l}}$ (the **short path**) along R and the short cuts of R from p to q.

4 Length of the Construction.

Let $L(P)$ be the length of the polygon P. The convex hull of P has length at most $L(P)$. Let $L(R)$ be the length of the boundary of a region R.

Let us first analyze the length of the segments added during the preliminary trimming. The length of the short cuts $p_0q_0, p_1q_1, \ldots$, (see section 2), form a decreasing geometric progression. Since the angle at the vertex v is at most $2\text{Arc}\sin 1/a$, it is easy to see that

$$\sum_{i\geq 0}|p_i - q_i| < \frac{1}{a-1}(|p_0 - v| + |v - q_0|).$$

After trimming the boundary, the boundary of the trimmed region R' has length $L(R') =$

$$\begin{aligned} &= L(R) + \textstyle\sum_{v \text{ trimmed}}(|p_0 - q_0| - |p_0 - v| - |v - q_0|) \\ &\leq L(R) + \textstyle\sum_{v \text{ trimmed}}(|p_0 - v| + |v - q_0|)(\frac{1}{a} - 1). \end{aligned}$$

Let R'_i be the region trimmed after adding the ith segment of the main construction, $R'_0 = R'$. Whenever we add the short cut pq, the boundary gets trimmed, and its length is reduced by at least $(a-1)|p-q|$. In the end it is still non-negative. Thus:

$$L(R') \geq (a-1)\sum_{\text{cuts } pq}|p - q|.$$

The total length of the short cuts added in region R is overall at most

$$\frac{1}{a-1}L(R') + \sum_{v \text{ trimmed}}\frac{1}{a-1}(|p_0 - v| + |v - q_0|)$$

$$\leq \frac{1}{a-1}(1 + \frac{1}{a})L(R).$$

As we sum over all the regions, every edge of P is counted at most twice, and every edge of the convex hull is counted only once. Thus the total length of the construction is at most

$$3\frac{a+1}{a(a-1)}L(P).$$

5 Analysis of Path Length.

Let p and q be two points of polygon P which are visible across region R. In section 3 we defined a path going from p to q. We must now show that this path γ really is a *short path* between p and q, i.e. its length is at most a constant times $|p - q|$.

The path γ can be decomposed into several parts: sections where the path follows the boundary of R, section where it hops across $V(p,q)$ (when $V(p,q)$ diverges from l with angle larger than α), and sections where it hops across other Voronoi circles.

In our analysis we will use the fact that $\tilde{l}$ is not too long compared to $|p - q|$.

Lemma 3 *The length of $\tilde{l}$ is at most*

$$|p - q|(1 + \sin\alpha)/\cos\alpha.$$

Proof: $\tilde{l}$ is a path from p to q, monotone in direction l, and can never diverge from l at angle steeper than α. □

5.1 Analysis of the boundary sections.

Consider a section where the path γ follows the boundary. Except near p and q, the corresponding section of $\tilde{l}$ goes between two successive intersections c_1 and c_2 of $\tilde{l}$ with the skeleton. Thus the boundary section is all within one skeleton cell, hence concave.

Case 1. The Voronoi circles of centers c_1 and c_2 intersect. The concave path must stay outside the circles, and so its length is at most equal to the distance from c_1 to c_2 (see figure 4).

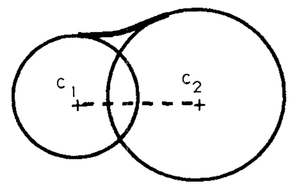

Figure 4: A concave path as shown cannot be longer than $|c_1 - c_2|$.

Thus the length of the boundary section is at most equal to the length of the part of $\tilde{l}$ between c_1 and c_2.

Case 2. The circles of centers c_1 and c_2 are disjoint.

Lemma 4 *In that case, the length of the boundary section is at most equal to*

$$2(1 + \frac{1}{\cos\alpha})|c_1 - c_2|.$$

Proof idea: The concave path must not intersect the section of $\tilde{l}$ between c_1 and c_2. The path $\tilde{l}$ is not necessarily a straight line between c_1 and c_2, but it is constrained with respect to direction l; thus $\tilde{l}$ has to lie inside a certain quadrilateral, and so does the concave path. Elementary geometric arguments (project the path outward to the boundary of the parallelogram) give the bound, see figure 5.

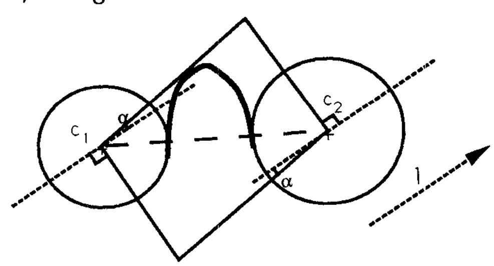

Figure 5: A concave path between separated circles must lie within a quadrilateral.

Case 3: Boundary sections near p and q. Initially, we follow the boundary of the region from p to the point on the first hopping circle.

Lemma 5 *This initial length is at most* $|p - c(p)|$.

Proof: Let p_1 be the first hopping point. The boundary between p and p_1 must be concave, and cannot intersect either l or the first hopping circle, see figure 6. The

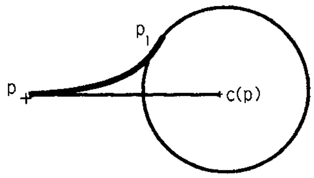

Figure 6: The initial section of γ is not too long.

section near q is dealt with similarly.

As we sum over all the cases, we find that the total length of the boundary sections of γ is at most

$$2(1 + \frac{1}{\cos\alpha})L(\tilde{l}).$$

5.2 Analysis of the hops

$\tilde{l}$ is a monotone path from p to q. We draw the sequence of centers of hopping points along $\tilde{l}$, along with the corresponding circles (see figure 8). The union of the circles form a region R.

Lemma 6 *The boundary of* R *has length* $L(\partial R) \leq 24L(\tilde{l})$.

Proof: Let $0 < \beta < 1$. We group the circles in batches. Let C_1 be the largest circle, with center c_1 and radius r_1. We associate all the circles before it and after it on the path $\tilde{l}$, which have center at a distance less than βr_1 away from the center of C_1, stopping at the first circle we meet (both before and after) which has center farther than βr_1 away. This defines the first batch of circles; the second batch is defined starting with the largest circle left, and so on until all the circles are in some batch.

We claim that the union of the circles in the ith batch has boundary length $\leq c(\beta)r_i$, where

$$c(\beta) = 2\pi\sqrt{\frac{1+\beta}{1-\beta}}$$

(this follows from elementary geometric considerations: the region is starred around c_i, has radius at most $(1+\beta)r_i$, and none of the tangents to the boundary is too steep with respect to c_i). Thus the boundary of R has length at most $c(\beta)\sum_{C_i} r_i$. On the other hand, we have

$$L(\tilde{l}) \geq \sum |c_i - c_{i+1}| \geq \sum_{\text{batches}} \beta r_j$$

Thus we have

$$L(\partial R) \leq \frac{c(\beta)}{\beta} L(\tilde{l}).$$

By adjusting β, we arrive at the constant stated. □

We now look at the path γ. For every hop between two points above $V(p,q)$, we map the hopping segment onto the arc of circle on its left, and for every hop between two points below $V(p,q)$, we map the hopping segment onto the arc of circle on its right. See figure 8. It is easy to argue that two different hops which do not cross $V(p,q)$ must have disjoint images under this map, and that the image is on the boundary of R. Thus the total length of the hops which are not across $V(p,q)$ is at most equal to the boundary of R, hence $\leq 24L(\tilde{l})$.

For the hops in the batch which do cross $V(p,q)$, we project them onto direction l; if α is not too small, in particular if

$$\sin(\alpha/2) > \beta/(1-\beta),$$

the projections of these hops are disjoint. Thus we have a contribution from these hops of

$$\sum_{\text{batches}} \frac{\text{diam(batch)}}{\sin\alpha} < \frac{2(1+\beta)L(\tilde{l})}{\beta\sin\alpha}.$$

Finally, if we consider all the contributions to γ, we find that the total length is $O(|p-q|)$, hence the theorem. The constants we obtain by minimizing over β and α are

$$C_1 = 77a \text{ when } C_0 = 3\frac{(a+1)}{a(a-1)}.$$

There is, however, a lot of room for more careful geometric analysis for the constant C_1, which is in practice much smaller.

6 Smooth curves and further generalizations

Having solved the short cut problem for polygons, we can solve it for smooth curves as follows.

Given a smooth curve S, define $\epsilon > 0$ so that any disk of diameter less than ϵ intersects S in a connected set, and in any disk of diameter ϵ centered on a point of S, the curve S does not deviate much from a straight line.

Take a polygon P approximating S to within $\epsilon/10$, such that the vertices of the polygon are not too acute, and let π be a map which is a one-to-one Lipschitz projection of the polygon onto the curve (in which π moves points by at most $\epsilon/10$, and the Lipschitz constant Lip_π is close to 1). We can find such a polygon and π since the curve S looks locally flat.

When we construct short cuts for the polygon P, there are no trimmed vertices, since the vertices of P are sufficiently obtuse. Given a short cut pq for P of length at least ϵ, we construct the short cut $\pi(p)\pi(q)$ on S. Note that this short cut has length at most 1.2 times the length of pq.

These are the only short cuts we use on S; their total length is at most 1.2 times the total length of the short cuts on P. We claim that they allow us to construct short paths for pairs of points on S.

To construct a short path from s_1 to s_2 if $|s_1 - s_2| < \epsilon$, simply follow the boundary, which is approximately a straight line from s_1 to s_2. If $|s_1 - s_2| > \epsilon$, first construct a short path from the corresponding points $\pi^{-1}(s_1)$ to $\pi^{-1}(s_2)$ on P; and then project the path using π to the set S and its short cuts. The length of the projection is at most $\max\{1.2, Lip_\pi\}$ times the length of the short path on P. □

The preceeding construction works for arbitrary rectifiable sets in the plane; these are the most general sets which are not locally 'twisted', which is the condition we need to make this approximation by polygonal shortcuts work.

References

[1] Peter Jones. *Rectifiable Sets and the Traveling Salesman Problem,* Invent. Math. 102, 1-15(1990).

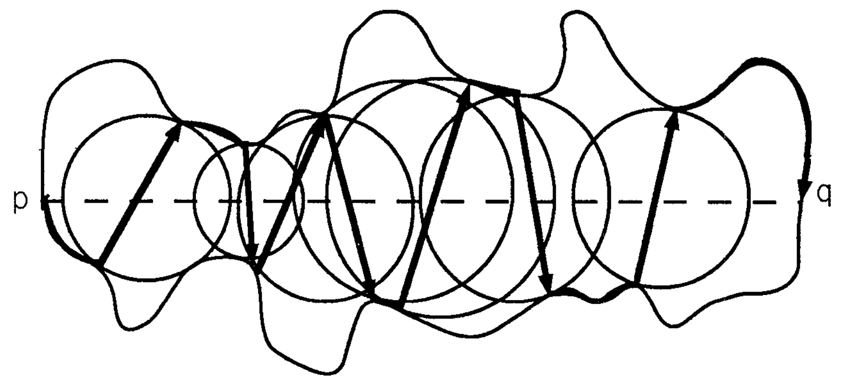

Figure 7: An inefficient path from p to q may be obtained by following l exactly.

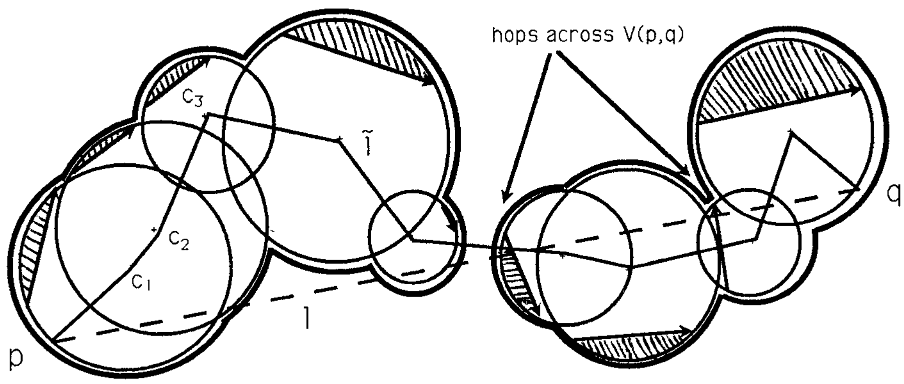

Figure 8: Dealing with the hops.