Shortest Paths in Line Arrangements

by

Anton Likhtarov

B.Sc., The University of British Columbia, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Computer Science)

The University of British Columbia
(Vancouver)

April 2020

© Anton Likhtarov, 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Shortest Paths in Line Arrangements

submitted by **Anton Likhtarov** in partial fulfillment of the requirements for the degree of **Master of Science** in **Computer Science**.

Examining Committee:

William Evans, Associate Professor, Department of Computer Science, UBC
Supervisor

David Kirkpatrick, Professor Emeritus, Department of Computer Science, UBC
Supervisory Committee Member

Abstract

The problem of finding a shortest Euclidean path in an arrangement of lines between two points in the arrangement has been extensively studied; however, the best known exact solution takes quadratic time, and it's not known if a subquadratic time algorithm exists. While I did not succeed in improving these bounds, I examined instead the problem of efficiently finding the approximate shortest path where the runtime depends on the bound of the relative error in the path length. I present an algorithm for computing this approximate shortest path. The algorithm uses the geometric structure of the arrangement; I show that certain lines are never used by the shortest path, while other lines could be ignored without making the path much longer. My work includes a number of lemmas that provide simple proofs for related problems (such as shortest path in two intersecting pencils of lines), and could have applications in future work on this problem.

Lay Summary

Imagine a city where every street is a straight line that extends to infinity in both directions. It is not known in general how quickly one can find the shortest possible route from one intersection to another in such a city.

I present a general method of finding a good enough route between two intersections that's in a certain sense guaranteed to not take too many calculations to find. I also present a number of new geometric theorems that will help to think about this problem in a new way, potentially leading to new discoveries in the future.

Preface

This thesis is original, unpublished, independent work by the author, Anton Likhtarov.

Table of Contents

List of Figures

Acknowledgments

I first would like to thank my advisor Will Evans for his never-ending patience and careful guidance, especially during the times when the progress seemed impossible.

I would also like to thank my second reviewer David Kirkpatrick for his valuable comments.

Also I thank my friends Ralph Furman and Reza Zadeh for reviewing the slides for my thesis presentation and for their good-humored ribbing about not having graduated yet over the years.

Finally, I thank my wife Stella—without her loving support and encouragement I might have never completed this thesis.

Chapter 1

Introduction

The shortest distance between two points is not a very interesting journey.
— Rube Goldberg

1.1 Shortest paths in line arrangements

We consider the problem of finding a shortest path (using the Euclidean metric) in an arrangement of lines on a plane between two points in the arrangement.

Given a finite set of lines $\mathscr{A} = \{l_i\}$ on a plane and two points S, T on some lines in $\mathscr{A}$, the SHORTEST-PATH problem is the problem of finding the shortest possible path p from S to T along the lines in $\mathscr{A}$.

Given also a real number $\varepsilon < 1$, the APPROXIMATE-SHORTEST-PATH problem is the problem of finding a path from S to T along the lines in $\mathscr{A}$ with length that is at most $(1+\varepsilon)$ times the length of p.

While these problems have been studied for some time, it's still an open problem to establish tight bounds on time complexity. The best known solution for SHORTEST-PATH has the time complexity $O(n^2)$, while the best known lower bound is $\Omega(n \log n)$.

While we didn't succeed in improving these bounds, our contributions are:

1. We show a new approach for exploring the shortest path geometry, proving the novel Circle Lemma, and exploring the connections to Urquhart's The-

orem. We use these techniques to demonstrate a simple, geometric proof of the Pencil Lemma (the existing proofs due to Kavitha [13] and Hart [11] involve full parameterization and complicated algebraic manipulations).

2. We present an algorithm for the APPROXIMATE-SHORTEST-PATH problem with the time complexity of $O(n \log n + \varepsilon^{-3} \log(\varepsilon^{-1})n)$.

1.2 History and related work

The earliest known treatment of this problem is by Davis in 1948 [4]. He shows that certain lines will never be used by the shortest path.

To obtain the naive upper bound, we could construct the entire $O(n^2)$ arrangement and then use Dijkstra's shortest path algorithm for an overall time complexity of $O(n^2 \log n)$. To improve on this, Henzinger et al. show how to construct the shortest path in planar graphs in $O(n)$ time, bringing the overall time complexity down to $O(n^2)$[12]. In addition, an approach called "topological peeling"[2, 3] reduces the space complexity to $O(n)$. As far as we can tell, this is the best currently known solution.

On the other hand, the best known lower bound of $\Omega(n \log n)$ follows from the reduction from the CONVEX-HULL-SIZE-VERIFICATION problem ("given n points on a plane, is their convex hull of size n?"), which is known to be $\Omega(n \log n)$ [15].

Some restricted versions of the problem have also been studied. If the lines are restricted to only k unique orientations, the shortest path can be found in time $O(n+k^2)$[6]. If the arrangement is formed by two intersecting pencils of lines and we're interested in the shortest path between the corners of the resulting "grid", the path is trivial and can be found in $O(n)$ time [11, 13]. It's interesting to note that even this seemingly much simpler variant of the problem has been open for some time, and when eventually solved both proofs involve full parameterization and complicated algebraic manipulations. One of our contributions is a simple geometric proof which we call the Pencil Lemma.

The APPROXIMATE-SHORTEST-PATH can be solved in $O(n \log n)$ if $\varepsilon = 1$ (that is, we want a factor of 2 approximation) [1].

Finally, Hart gives a short overview for an approximation algorithm with time complexity of $O(n\log n + (\min\{n, \varepsilon^{-2}\})\varepsilon^{-1}\log(\varepsilon^{-1}))$[10], though the complete algorithm and proof was never published. We build on some of the ideas in this overview paper and present a complete approximation algorithm.

Chapter 2

Exploring the arrangement geometry

2.1 Definitions and observations

We assume that no line contains both S and T; otherwise the shortest path is trivially found in $O(n)$ time.

We fix a coordinate system so that $S = (-1, 0)$ and $T = (1, 0)$.

We use the notation $d(A,B)$ for the length of the segment $\overline{AB}$, and $d(A,B,C)$ as a shortcut for $d(A,B) + d(B,C)$ (similarly for any larger number of terms).

Every line l induces a partition on the plane $l \cup l^+ \cup l^-$ where l^+ and l^- are the open half-planes defined by l. A line l that separates S from T (that is, S and T are in different half-planes defined by l) is called a *cross line*; every other line is called an *exterior line*.

For every exterior line b, take the closed half-plane defined by b that contains both S and T. The intersection of these half-planes is a convex, possibly unbounded region that we denote $\text{Hull}(S,T)$. An exterior line that has a nonempty intersection with $\text{Hull}(S,T)$ is called a *boundary line*.

It's easy to see that if a shortest path intersects a line, the intersection is a closed segment or a point. The shortest path will intersect every cross line, and will never leave $\text{Hull}(S,T)$. Thus we can ignore any non-boundary exterior lines.

A path is said to *visit* a line if it intersects with it. A path *travels* on a line if the intersection is a positive length segment and not just a point.

We will often need to explicitly state the direction that a path is travelling along a given line. It's trivial to see that a shortest path will only travel on lines on the upper boundary of $\mathrm{Hull}(S,T)$ in the clockwise direction from S to T—otherwise the path would self-intersect, and we could obtain a shorter path. Similarly, a shortest path will only travel on lines on the lower boundary in the counterclockwise direction from S to T.

The shortest path can travel on a cross line in either direction. To make the discussion clear, we replace each cross line with two coincident directed lines: an *up-line* with the positive y-direction of travel and a *down-line* with the negative y-direction of travel.

Now that we have partitioned the arrangement into a set of upper boundary, lower boundary, up-, and down-lines, the possible direction of travel on each line is unique.

Observation 2.1.1. *If a shortest path travels consecutively on lines x and then y,*

1. *x is an upper boundary or an up-line if and only if y is an upper boundary or a down-line.*

2. *x is a lower boundary or a down-line if and only if y is a lower boundary or an up-line.*

Proof. Sketch. In each case, the negation would have the path cross some line twice. For example, if x and y were both down-lines, then p would either have to cross x again after travelling on y or would have already crossed y before travelling on x. □

Finally, we have some notation to help us talk about line angles. An up-line a is said to be *steeper* than an up-line b if the angle that the directed line a makes with the directed line $\overrightarrow{ST}$ is larger than the angle that b makes with $\overrightarrow{ST}$. We write $b < a$ (b is *less steep* or *shallower* than a). We have a similar definition for down-lines. Intuitively, the path would often prefer a shallower line to get across $\mathrm{Hull}(S,T)$ faster.

2.2 General position

For the simplicity of the argument, we assume that the arrangement given is in general position. By this we mean that no two lines are parallel; no three lines intersect at a single point; and S and T belong to exactly one line.

Note that any arrangement can be perturbed slightly to be in general position without changing the sequence of lines visited by a shortest path (with perhaps the exception of adding a small segment of travel next to S or T if they happen to intersect more than one line). Unfortunately it is not known if we can compute this perturbation quickly—the problem of finding if any three lines intersect at a point belongs to a class of "n^2-hard problems"[7, 14].

We believe that the argument can be adapted to drop the general position requirement, but we did not attempt to do so.

2.3 Urquhart's theorem and related lemmas

The following curious theorem has direct application to the problems we're considering.

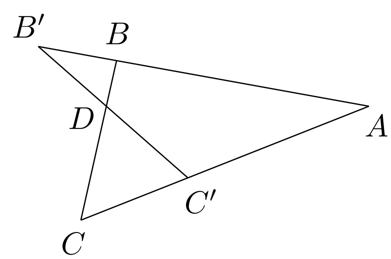


Figure 2.1: Urquhart's theorem.

Theorem 2.3.1 (Stronger version of Urquhart's Theorem [8]). *If $\overleftrightarrow{ABB'}$[1] and $\overleftrightarrow{AC'C}$ are straight lines with $\overleftrightarrow{BC}$ and $\overleftrightarrow{B'C'}$ intersecting at D then*

1. $d(A,B,D) = d(A,C',D)$ *if and only if* $d(A,B',D) = d(A,C,D)$*, and*

2. $d(A,B,D) < d(A,C',D)$ *if and only if* $d(A,B',D) < d(A,C,D)$.

[1] We use $\overleftrightarrow{AB}$ for the line through the points A and B; $\overrightarrow{AB}$ for the directed line through A and B with the direction of travel from A towards B; and $\overline{AB}$ for the closed line segment from A to B.

Pedoe attributes the theorem to L. M. Urquhart who "discovered it when considering some of the fundamental concepts of the theory of special relativity" [17], though it appears that the history of the theorem and its proofs go back much further [5, 16]. Hajja has published two proofs of the theorem with some discussion of its history [8, 9].

2.3.1 Pencil Lemma

The SHORTEST-PATH problem remained open even for the special case where the arrangement is formed by two intersecting pencils of lines (see Figure 2.2) until solved by Kavitha [13] and independently by Hart in 2003 [11].

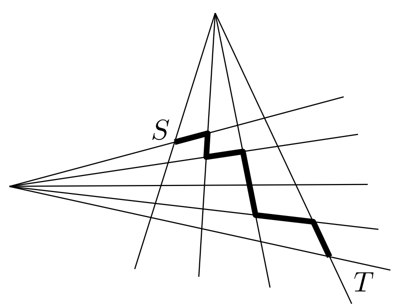


Figure 2.2: An *S*–*T* path on two intersecting pencils of lines. The shortest path would not use interior intersections.

It turns out that the shortest path never uses the interior of the "distorted grid" in this case, and it's sufficient to prove the case for the intersecting pencils of two and three lines. We refer to this base case as the Pencil Lemma.

We're aware of two existing proofs. Kavitha [13] parameterizes the lengths of certain paths through the coordinates of the line intersections and differentiates the resulting functions to show that certain path length relationships must hold. Hart's approach [11] is similar but uses angles and trigonometry for the parameterization instead; it also uses Mathematica to simplify the resulting expressions. Neither approach seems to provide any geometric insight into the problem.

We present a simpler geometric proof that only uses Urquhart's Theorem and the triangle inequality.

Lemma 2.3.2 (Nested Ellipse Lemma)**.** *For $\overleftrightarrow{PXY}$ a straight line with $X \neq Y$ and $Q \notin \overleftrightarrow{PXY}$, if $d(Z,Q,X) \leq d(Z,P,X)$ then $d(Z,Q,Y) < d(Z,P,Y)$.*

Alternatively, the ellipse with foci Z and X is contained within the ellipse with foci Z and Y if the ellipses coincide at P (see Figure 2.3).

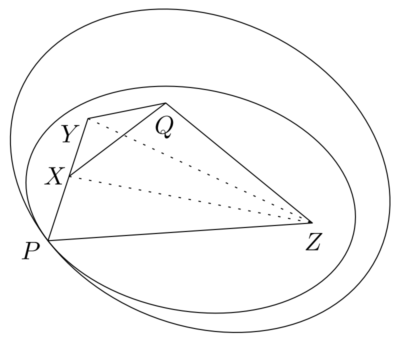


Figure 2.3: Nested Ellipse Lemma.

Proof.

$$\begin{aligned} d(Z,P,Y) &= d(Z,P,X,Y) \\ &\geq d(Z,Q,X,Y) \\ &> d(Z,Q,Y) \qquad \text{by the triangle inequality.} \end{aligned}$$

□

Lemma 2.3.3 (Pencil Lemma)**.** *Fix points S and T. For all points P, Q, and $U \in \overleftrightarrow{PS}$, a shortest S–T path on the arrangement $\{\overleftrightarrow{PS}, \overleftrightarrow{PT}, \overleftrightarrow{QS}, \overleftrightarrow{QT}, \overleftrightarrow{QU}\}$ does not use the segment $\overline{UV}$, where $V = \overleftrightarrow{PT} \cap \overleftrightarrow{QU}$, provided that no two lines are parallel.*

Proof. Let $A = \overleftrightarrow{PS} \cap \overleftrightarrow{QT}$ and $B = \overleftrightarrow{PT} \cap \overleftrightarrow{QS}$.

If $U \notin \overline{SA}$ then the path that uses $\overline{UV}$ would double-cross a line, and the result holds, so assume $U \in \overline{SA}$. The three possible paths from S to T have lengths $d(S,A,T)$, $d(S,B,T)$, and $d(S,U,V,T)$.

We show that $\min\{d(S,A,T), d(S,B,T)\} < d(S,U,V,T)$. It suffices to show that if $d(S,U,V,T) \leq d(S,B,T)$ then $d(S,A,T) < d(S,U,V,T)$, or equivalently if $d(S,U,V) \leq d(S,B,V)$ then $d(U,A,T) < d(U,V,T)$—in other words, if the path

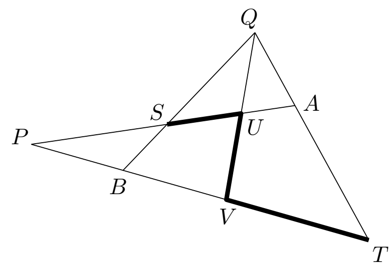


Figure 2.4: Pencil Lemma. The path shown cannot be shortest in the arrangement.

that uses $\overline{UV}$ is better than the path through B then the path through A must be better still.

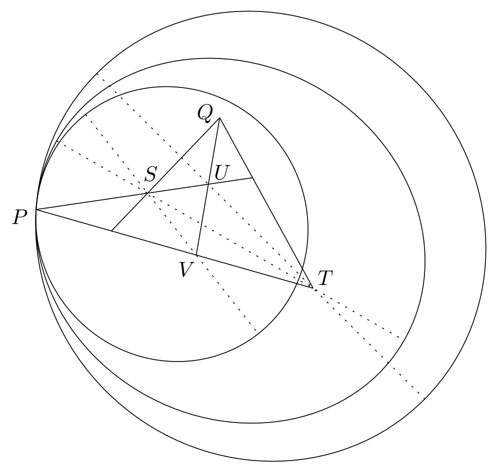


Figure 2.5: Three nested ellipses through P. The ellipse with foci at S and V is inside the ellipse with foci at S and T, which is inside the ellipse with foci at T and U.

Assume $d(S,U,V) \leq d(S,B,V)$.

By Theorem 2.3.1, $d(S,Q,V) \leq d(S,P,V)$.

By Lemma 2.3.2, taking $X = V$, $Y = T$, and $Z = S$, $d(S,Q,T) < d(S,P,T)$.

By Lemma 2.3.2 again, taking $X = S$, $Y = U$, and $Z = T$, $d(U,Q,T) < d(U,P,T)$.

Finally, by Theorem 2.3.1, $d(U,A,T) < d(U,V,T)$.

Intuitively, there are three nested ellipses through P (see Figure 2.5), so it follows that if Q is inside the innermost ellipse, it must also be inside the outermost ellipse and the result follows. $\square$

2.3.2 Quadrilateral Lemma

We'll make use of the following simple geometric lemma, which is closely related to Urquhart's Theorem. Using this lemma, we show that a given shortest path cannot travel on some line segments in the arrangement since there would be a way to shorten the path.

Lemma 2.3.4. *Let a, b, and c be three lines tangent to a circle $\mathscr{C}$ so that $b \cap c$ and $\mathscr{C}$ are in different half-planes defined by a.*

Let $P = a \cap b$, $Q = b \cap c$, and $R = a \cap c$.

Let x be a line so that $U = x \cap a \in \overline{PR}$ and $V = x \cap c \in \overline{RQ}$.

1. *If x is tangent to $\mathscr{C}$ then $d(U,P,Q) = d(U,V,Q)$.*
2. *If x crosses $\mathscr{C}$ then $d(U,P,Q) > d(U,V,Q)$.*
3. *If x does not cross $\mathscr{C}$ then $d(U,P,Q) < d(U,V,Q)$.*

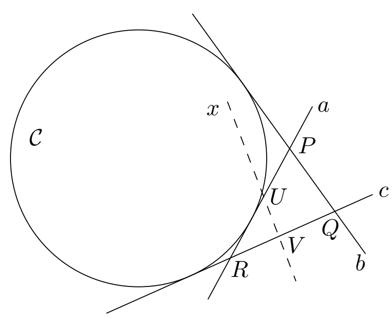


Figure 2.6: Lemma 2.3.4.

Proof. **Case 1.** x is tangent to $\mathscr{C}$.

Let A, B, C, X be the points where a, b, c, and x respectively coincide with $\mathscr{C}$ (see Figure 2.7).

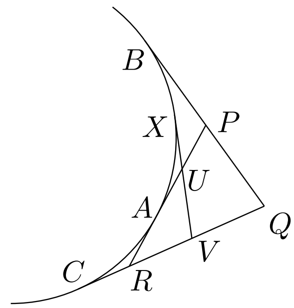


Figure 2.7: Case 1: x is tangent to $\mathscr{C}$.

We have:

$$d(C,Q) = d(B,Q) \qquad d(C,V) = d(X,V)$$
$$d(A,P) = d(B,P) \qquad d(A,U) = d(X,U)$$
$$d(C,Q) = d(C,V,Q) = d(X,V,Q) = d(X,U,V,Q) = d(A,U,V,Q)$$
$$d(B,Q) = d(B,P,Q) = d(A,P,Q) = d(A,U,P,Q)$$

So $d(A,U,V,Q) = d(A,U,P,Q)$, and thus $d(U,V,Q) = d(U,P,Q)$. □

Case 2. x crosses $\mathscr{C}$ (see Figure 2.8).

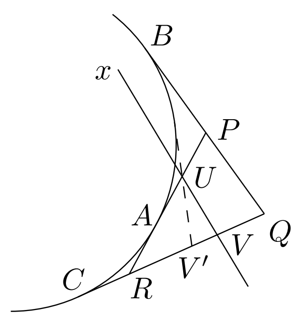


Figure 2.8: Case 2: x crosses $\mathscr{C}$.

Let $\overleftrightarrow{UV'}$ be the line through U tangent to $\mathscr{C}$ that does not coincide with a. By Case 1, $d(U,P,Q) = d(U,V',Q) = d(U,V',V,Q) > d(U,V,Q)$ (by the triangle inequality). □

Case 3. x does not cross $\mathscr{C}$.

As in Case 2, Let $\overleftrightarrow{UV'}$ be the line through U tangent to $\mathscr{C}$ that does not coincide with a. By Case 1, $d(U,P,Q) = d(U,V',Q) < d(U,V,V',Q) = d(U,V,Q)$ (by the triangle inequality). $\square$

2.4 Circle Lemma

The Circle Lemma shows that travelling along any cross line imposes significant restrictions on any other lines crossed by the shortest path.

Lemma 2.4.1 (Circle Lemma)**.** *Let $\mathscr{C}$ be the unique circle tangent to lines a, b, and c that lies in the half-plane defined by a that does not contain $b \cap c$.*

If b separates S from T and the shortest path from S to T travels consecutively on a, b, and c, then the arrangement contains no lines that intersect both $\mathscr{C}$ and the shortest path from S to T.

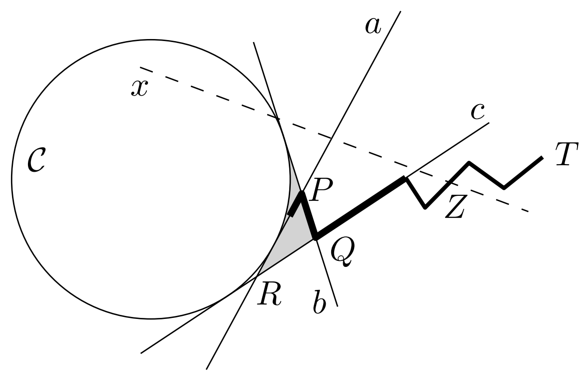


Figure 2.9: Circle Lemma. No line intersects both $\mathscr{C}$ and the shortest path. S is not shown, but is somewhere within the shaded area.

Proof. We first establish some common notation and then prove the lemma in multiple parts.

Let $P = a \cap b$, $Q = b \cap c$, $R = a \cap c$. Suppose for the sake of contradiction that there's a line x that intersects both $\mathscr{C}$ and the shortest path from S to T at a point Z (see Figure 2.9).

Part 1. *Z is located in the shortest subpath from S to Q.*

We first observe that Z must be inside the same quadrant defined by the lines b and c that contains the circle $\mathscr{C}$; in all other cases the supposed shortest path from S to Q would double cross some line which is a contradiction.

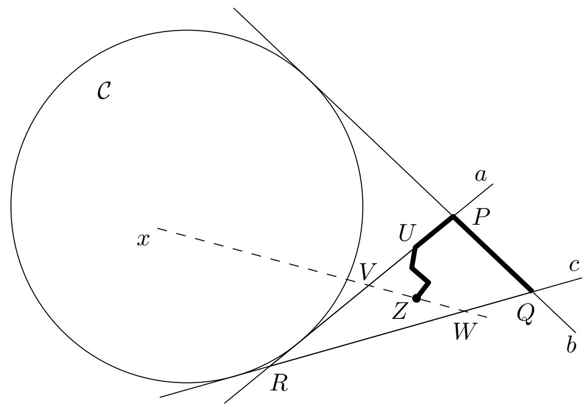


Figure 2.10: First case: $Z \in \triangle PQR$.

First, consider the case where $Z \in \triangle PQR$ (see Figure 2.10).

Let $V = a \cap x$ and $W = c \cap x$. It must be that $W \in \overline{RQ}$, otherwise x would have crossed the shortest path twice, which is a contradiction. It follows that $V \in \overline{RP}$ since x crosses $\mathscr{C}$.

If δ is the length of the shortest subpath from Z to Q, we have

$$\begin{aligned} d(Q,W,V) < d(Q,P,V) &\leq \delta + d(Z,V) \quad \text{(by Lemma 2.3.4)} \\ d(Q,W,Z,V) = d(Q,W,V) &< \delta + d(Z,V) \\ d(Q,W,Z) &< \delta, \end{aligned}$$

which is a contradiction.

The other case is where $Z \notin \triangle PQR$ (see Figure 2.11).

Let d be the line on which the shortest path arrives at U, and $V \in d$ be some point so that $\overline{VU}$ is part of the shortest path.

Let $W = d \cap c$. It must be that $W \in \overline{RQ}$, otherwise d would have crosses the shortest path twice.

Note also that d does not cross $\mathscr{C}$ since otherwise $d(U,W,Q) < d(U,P,Q)$ (by Lemma 2.3.4), which would be a contradiction.

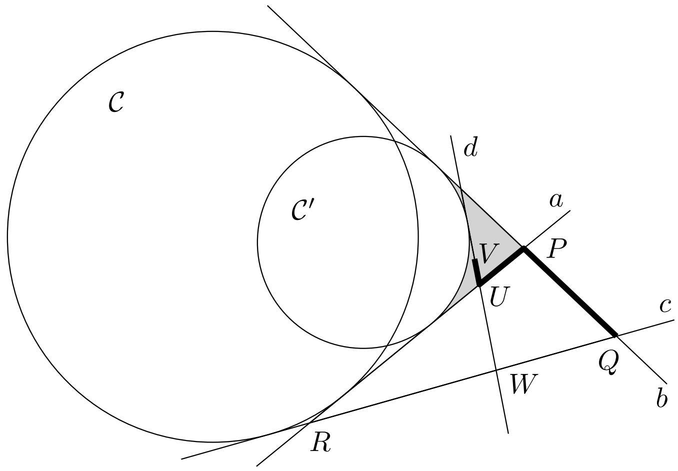


Figure 2.11: Second case: $Z \notin \triangle PQR$. Z is located within the shaded area.

We apply the Circle Lemma by induction on the number of line segments in the shortest path, considering the shortest path from $S' = Z$ to $T' = Q$, and taking $a' = d$, $b' = a$ (which separates Z from Q), and $c' = b$. Since the shortest path from Z to Q has at least one segment fewer than the shortest path from S to T and is finite, the induction must eventually end in the first case of the proof.

Let $\mathscr{C}'$ be the circle in the inductive step. We have that no line through Z crosses $\mathscr{C}'$. As a corollary, Z is located in the bounded region defined by the lines a, b, and the circle $\mathscr{C}'$.

Any line through Z that crosses $\mathscr{C}$ must also cross $\mathscr{C}'$, since $\mathscr{C}$ and $\mathscr{C}'$ share the tangents a and b, and $\mathscr{C}$ contacts these tangents at points that are further away from P than $\mathscr{C}'$ does (by virtue of another tangent d of $\mathscr{C}'$ not crossing $\mathscr{C}$).

So no lines through Z can cross $\mathscr{C}$.

Part 2. *Z is located in the shortest subpath from Q to T.*

Since b separates S from T and the shortest path cannot cross a or b twice, Z must be in the quadrant defined by a and b that is opposite of the quadrant containing $\mathscr{C}$.

Consider first the case where $Z \in c$ or Z is in the same half-plane defined by c as the circle $\mathscr{C}$ (see Figure 2.12).

Let $W = x \cap a$. Note that $W \notin \overline{RP}$, otherwise x would cross $\overline{PQ}$, which is part

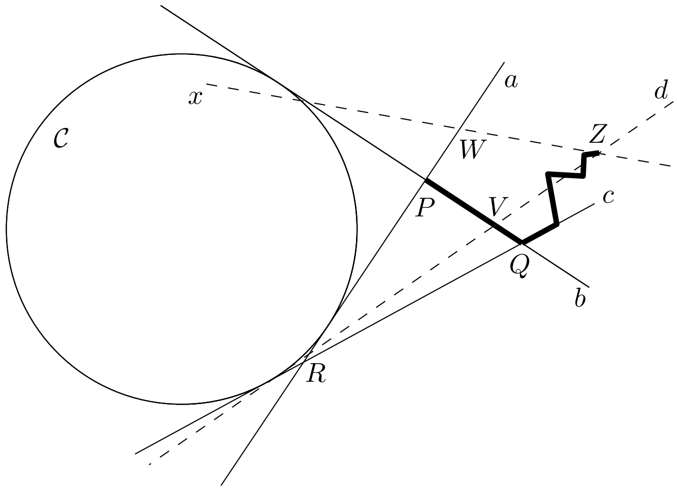


Figure 2.12: Case one: Z above c.

of the shortest path, a contradiction. Let d be the unique tangent line of $\mathscr{C}$ through Z that intersects $\overline{PQ}$, and let $V = d \cap \overline{PQ}$ (d might not exist in the arrangement).

Let δ be the length of the shortest subpath from P to Z. By Lemma 2.3.4, $d(P,W,Z) < d(P,V,Z) \leq \delta$, which is a contradiction.

The other case is where Z is in the half-plane defined by c that does not contain the circle $\mathscr{C}$. Let d be the line on which the path leaves c. We have two subcases.

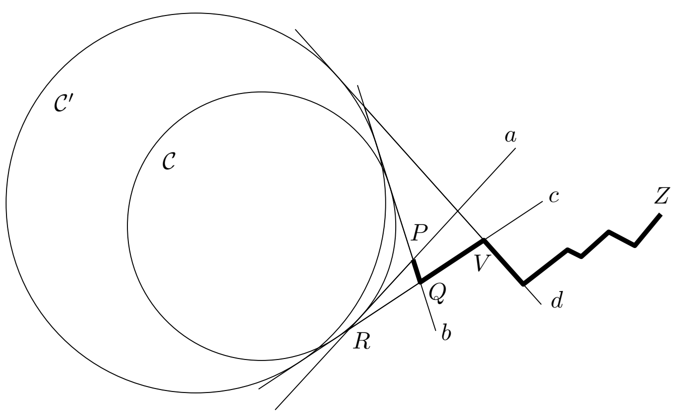


Figure 2.13: Case two: d intersects b above, induction.

First, consider the case where $d \cap b$ is in the same half-plane defined by c as the circle $\mathscr{C}$ (see Figure 2.13). Note that d does not cross $\mathscr{C}$ since otherwise we would have (by Lemma 2.3.4) $d(P, a\cap d, V) < d(P, Q, V)$, a contradiction.

We apply the Circle Lemma by induction on the number of line segments in the shortest path, considering the shortest path from $S' = P$ to $T' = Z$, and taking $a' = b$, $b' = c$ (which separates P from Z), and $c' = d$. Since the shortest path from P to Z has at least one segment fewer than the shortest path from S to T and is finite, the induction must eventually end in one of the other cases of the proof.

Let $\mathscr{C}'$ be the circle in the inductive step. We have that no line through Z crosses $\mathscr{C}'$.

Since Z is in the quadrant defined by b and c that is opposite to the quadrant containing $\mathscr{C}$ and $\mathscr{C}'$, any line through Z that crosses $\mathscr{C}$ must also cross $\mathscr{C}'$, because $\mathscr{C}$ and $\mathscr{C}'$ share the tangents b and c, and $\mathscr{C}'$ contacts these tangents at points that are further away from Q than $\mathscr{C}$ does (by virtue of another tangent d of $\mathscr{C}'$ not crossing $\mathscr{C}$).

So no lines through Z cross $\mathscr{C}$ in this case.

For the final case we have $d \cap b$ in the half-plane defined by c that does not contain the circle $\mathscr{C}$ (see Figure 2.14).

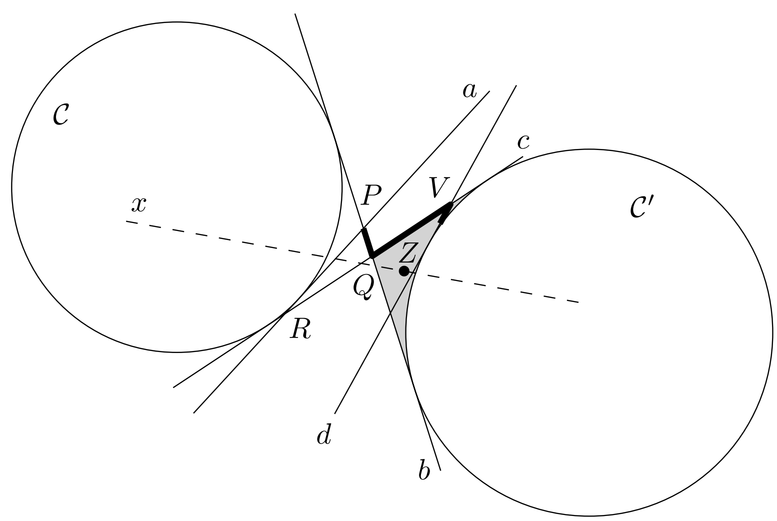


Figure 2.14: Case three: d intersects b below. Z can only appear in the shaded region, any line through Z that crosses $\mathscr{C}$ would also cross $\mathscr{C}'$.

We again apply the Circle Lemma by induction, considering the reversed shortest path from Z to P, and taking $a' = d$, $b' = c$ (which separates P from Z), and $c' = b$. In this case, we're guaranteed to end up in the Part 1 of the proof.

We get that in the inductive step no lines through Z cross the unique circle $\mathscr{C}'$ tangent to lines d, c, and b that lies in the half-plane defined by d that does not contain Q. We also get that the subpath from P to Z does not cross $\mathscr{C}'$, so Z must be in the bounded region defined by the lines b, c, and the circle $\mathscr{C}'$. Since b and c are tangent to both $\mathscr{C}$ and $\mathscr{C}'$, it follows that any line through Z that crosses $\mathscr{C}$ must also cross $\mathscr{C}'$.

Therefore no lines through Z cross $\mathscr{C}$ and this completes the proof. □

2.5 Phantom-line Lemma

This lemma allows us to conclude that the shortest path does not travel on certain line segments that are surrounded by "better" lines on each side.

Lemma 2.5.1 (Phantom-line Lemma)**.** *Let a, b, and c be three directed lines so that the shortest path from S to T travels through points $A \in a$ and $C \in c$ so that C is to the right of a and A is to the left of c.*

If b separates S from T and A from C, and b makes a positive angle with both a and c then the shortest path does not travel along b in the direction of the line (note: the shortest path may travel along a or c with or against the respective directions).

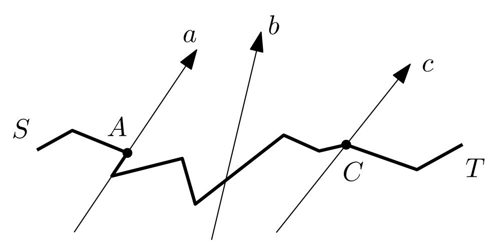


Figure 2.15: Phantom-line Lemma: the shortest path does not travel along b in its direction.

Proof. Suppose that the shortest path p does travel along a line segment $\overline{XY} \in b$.

$\overline{XY}$ must be to the right of a and to the left of c in its entirety, otherwise the path would double cross either a or c.

Let x be the line on which the shortest path arrives at X and y be the line on which the shortest path leaves Y.

Since b separates S from T, the shortest path travels along x and y on opposite sides of b. We will introduce a new line b' to the arrangement which is initially coincident with b (the “phantom line” of the lemma’s name). We will translate b' in steps, always maintaining (1) b' separates S from T and A from C; (2) using b' makes for a shorter path; and (3) after every step the shortest path (with b') has fewer segments in it, allowing us to use induction.

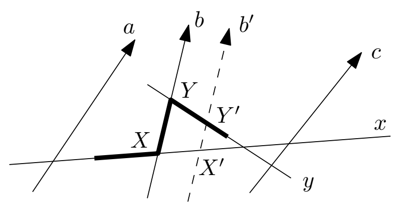


Figure 2.16: Translating the phantom line b'.

We continuously translate b' towards $x \cap y$. Note that as long as b' doesn’t cross $x \cap y$ or any vertices of the path, using b' instead of b results in a shorter path. Let $X' = x \cap b'$ and $Y' = y \cap b'$ (see Figure 2.16).

One of the following will happen as we’re translating b':

1. b' intersects a vertex Z on the shortest path (see Figure 2.17). If the path visits Z after leaving $\overline{X'Y'}$, we can obtain a shorter path p' by replacing the subpath from X' to Z with $\overline{X'Z}$ and using $x \cap z$ as the new direction of translation for b', where z is the line on which the path leaves Z. Similarly, if Z is visited before arriving at $\overline{X'Y'}$, we replace the subpath from Z to Y' with $\overline{ZY'}$ and translate towards $z \cap y$. The new path has fewer line segments, and the argument must eventually terminate in one of the other cases. (Note: we cannot use a more straightforward induction. While applying the lemma by induction tells us that there’s a shorter path in the new arrangement that does not travel along b' in the direction of the line, it could travel along b' in the *opposite* direction, which would not allow us to conclude that the same shortest path exists in

the original arrangement.)

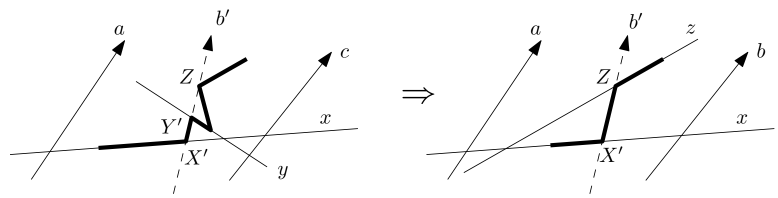


Figure 2.17: b' intersects Z. Note that the direction of translation can change: $z \cap x$ is on the opposite side of b' compared to $x \cap y$.

2. b' intersects $x \cap y$. In this case travel on b' has shrunk to a single point. We obtained a shorter path that does not rely on b', which is a contradiction.

3. $\overline{X'Y'}$ intersects a (see Figure 2.18). Since b (and thus b') makes a positive angle with a, we must have $Y' \in a$ happen first as we're translating b'. We replace the path from A to Y' with $\overline{AY'}$ to obtain a shorter path that does not use b', a contradiction.

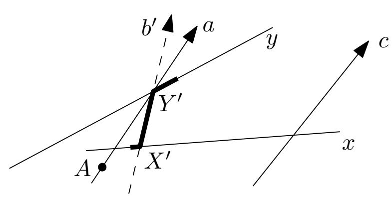


Figure 2.18: Y' lies on a.

4. $\overline{X'Y'}$ intersects c (see Figure 2.19). Since b (and thus b') makes a positive angle with c, we must have $X' \in c$ happen first as we're translating b'. We replace the path from X' to C with $\overline{X'C}$ to obtain a shorter path that does not use b', a contradiction.

$\square$

2.6 Two-lines Lemma

The following lemma allows us to only consider arrangements where no two up-lines (or down-lines) intersect within $\text{Hull}(S,T)$.

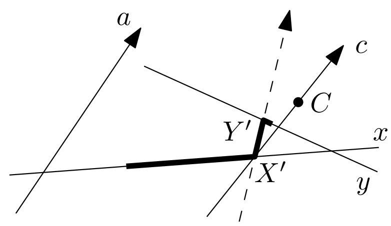


Figure 2.19: X' lies on c.

Lemma 2.6.1 (Two-lines Lemma)**.** *If $a < b$ are two up-lines (down-lines) that intersect within* $\mathrm{Hull}(S,T)$*, then no shortest path travels on b.*

Note that the condition that a and b intersect within $\mathrm{Hull}(S,T)$ is required since it is possible to construct a counterexample if $a \cap b \notin \mathrm{Hull}(S,T)$ (see Figure 2.20).

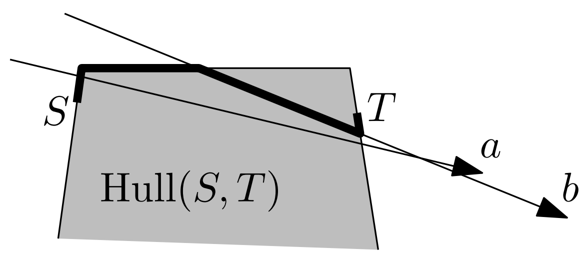


Figure 2.20: A counterexample where $a \cap b \notin \mathrm{Hull}(S,T)$.

Proof. We will prove the lemma for down-lines; the proof for up-lines is similar.

Suppose, for the sake of contradiction, that there is a shortest path p that travels on b. Let $\overline{BE}$ be the segment of travel on b. We must have $a \cap b \notin \overline{BE}$, otherwise the path would double-cross a, so there are two possibilities: either p visits a before $\overline{BE}$—that is, $\overline{BE}$ and T are in the same half-plane defined by a (see Figure 2.21); or p visits a after $\overline{BE}$—that is, $\overline{BE}$ and S are in the same half-plane defined by a.

In the second case, consider the rotation of the arrangement through π, with S and T swapped; the direction of travel along p is reversed and a and b are down-lines in the rotated arrangement. p visits a before $\overline{BE}$ in the rotated arrangement, so we have reduced this case to the first one (see Figure 2.22).

So we assume that p visits a before $\overline{BE}$.

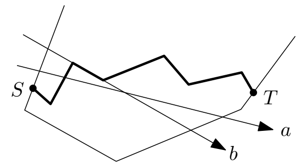
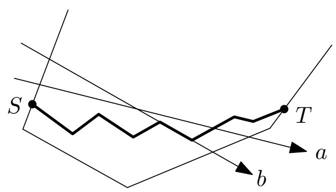


Figure 2.21: p visits a before travelling on b.



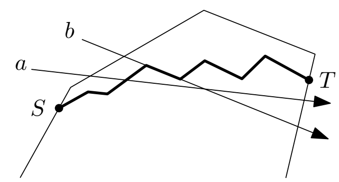


Figure 2.22: If p travels on b before visiting a then rotating the arrangement and reversing the shortest path direction reduces to the first case.

We further assume that b is the last down-line that p travels on so that b intersects one or more down-lines a' where $a' < b$, $a' \cap b \in \text{Hull}(S,T)$, and p visits a' before travelling on b, and we let a be the last a' that p travels on. Let $Q = a \cap b \in \text{Hull}(S,T)$, and let A be the point at which p leaves a.

Let x be the line on which p travels after leaving b. Note that since $E = x \cap b \in \overline{BQ}$, x cannot be a boundary line, otherwise $Q \notin \text{Hull}(S,T)$, so x is an up-line.

Let y be the line on which p travels after leaving x. Note that y must exist: if p arrives directly to T on x, then T is at an intersection of x and the boundary, which we disallow by the general position assumption.

Finally, let z be the line on which the path arrives to b (so that $z \cap b = B$). Let $e = \overrightarrow{AB}$ (which might not exist in the arrangement). Let further $C = e \cap y$, $D = a \cap x$, $F = x \cap y$, $Z = z \cap y$.

Consider first the case where $z \leq e$.

Claim 1. $y \cap a \in \overline{DQ}$.

Proof. Let $y' = \overleftrightarrow{FQ}$, $C' = e \cap y'$, and $Z' = z \cap y'$ (see Figure 2.23).

By Lemma 2.3.3 we have

$$\min\{d(A,D,F), d(A,C',F)\} < d(A,B,E,F)$$

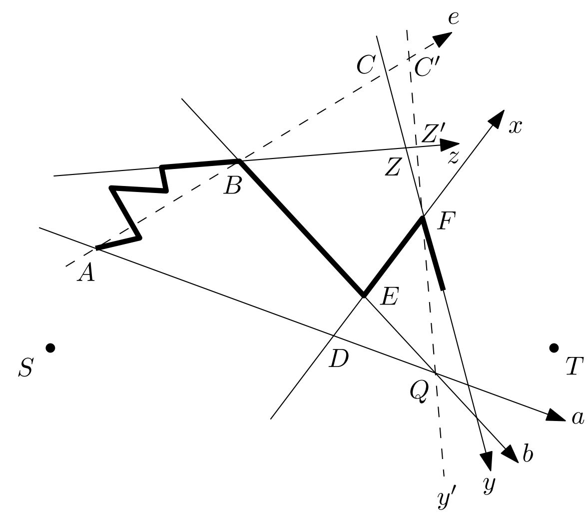


Figure 2.23: First case: $z \leq e$. $y \cap a \notin \overline{DQ}$ leads to a contradiction.

Note also that $d(A,B,E,F) \leq \delta$ where δ is the length of the actual subpath from A to F.

It can't be that $d(A,D,F) < d(A,B,E,F)$ since then shortest path would have travelled on $\overline{AD}$ and $\overline{DF}$, so we must have $d(A,C',F) < d(A,B,E,F)$ and thus $d(B,C',F) < d(B,E,F)$.

Since by assumption $z \leq e$, we also have $d(B,Z',F) \leq d(B,C',F)$.

If $y \cap a \notin \overline{DQ}$ then also $d(B,Z,F) < d(B,Z',F)$, so the path travelling on $\overline{BZ}$ and $\overline{ZF}$ would be shorter than p which travels on $\overline{BE}$ and $\overline{EF}$, a contradiction. $\square$

By Observation 2.1.1, since x is an up-line, y is either an upper boundary or a down-line. Since $y \cap a \in \overline{DQ}$ and $Q \in \text{Hull}(S,T)$, y cannot be a boundary line. So y must be a down-line.

Note that we have p travelling on y after visiting b, $b < y$, and $b \cap y \in \text{Hull}(S,T)$. This contradicts the assumption we made that b was the last line with these properties visited by p.

Consider now the case where $z > e$ (see Figure 2.24). Let w be the line on which p travels before arriving to z. Note that $w \neq a$, otherwise we would have $z = e$.

By Observation 2.1.1, and the fact that the travel on both w and z intersects the

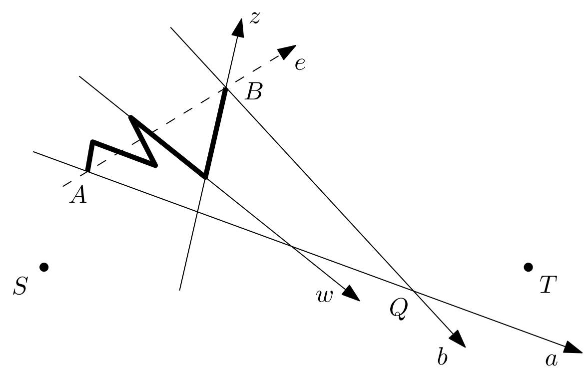


Figure 2.24: Second case: $z > e$.

interior of $\triangle ABQ$, we conclude that w is a down-line and z is an up-line; otherwise, if z is an upper boundary line then $A \notin \text{Hull}(S,T)$, and if w is a lower boundary line then either $A \notin \text{Hull}(S,T)$ or $Q \notin \text{Hull}(S,T)$.

Note also that we must have $w \cap a \in \overline{AQ}$ and $w \cap b \notin \overline{BQ}$—otherwise we would have $w \cap b \in \text{Hull}(S,T)$ and $w < b$, and we would have chosen w over a since the path travels on w after visiting a.

Consider the shortest subpath from S to B, and note that $w \cap a \in \text{Hull}(S,B)$. By induction on the number of segments in the shortest path, we have that the subpath would not travel on w, a contradiction. $\square$

Chapter 3

The $(1+\varepsilon)$ approximation algorithm

3.1 Overview

We present a high level summary of the algorithm steps here. The algorithm works by ignoring certain lines and intersections for the purpose of finding the shortest path. The resulting graph is small enough to make a dent in the time complexity, while at the same time we take care to ignore lines and intersections that could have made the shortest path only moderately shorter.

The approach borrows ideas from the overview paper by Hart [10], in particular the overall high level approach of constructing the unimodal sequence of angles and the insight that further elimination of lines by making the angles get exponentially small allows for an asymptotic reduction in running time while keeping the error in the path small. The full algorithm was never published so we don't know how Hart tackled some of the issues we cover here (in particular the approximation on the critical triangle). As far as we know the Circle Lemma and its application here is novel.

The subsequent sections fill in the details, prove that the algorithm is correct (that is, if the shortest path uses the lines or intersections that the algorithm ignores, it would be shorter by a factor of at most $(1+\varepsilon)$), and show that the time complexity is $O(n\log n+\varepsilon^{-3}\log(\varepsilon^{-1})n)$.

1. Partition the arrangement $\mathscr{A}$ into the set E of exterior lines and the set C of cross lines (that is, the lines that cross $\overline{ST}$).

2. Construct the set of up-lines U and the set of down-lines D by constructing the two directed lines for every line in C.

3. Construct $\mathrm{Hull}(S,T)$ by intersecting the half-planes defined by the lines in E (this can be done e.g. using the CONVEX-HULL algorithm in the dual plane). Construct two ordered sets of boundary lines B_U (the *upper boundary*) and B_L (the *lower boundary*) by taking those lines in E that define the upper and the lower chains of $\mathrm{Hull}(S,T)$ respectively.

4. Remove certain cross lines from U and D by constructing the simple subarrangement (see Section 3.2).

5. Remove cross line to cross line intersections within:

   upper and lower quadrants (see Section 3.5);

   the exponential subarrangement with $\varepsilon' = \sqrt[3]{1+\varepsilon} - 1$ (see Section 3.6.1);

   and the critical S and T triangles, taking the same ε' as above for both (see Section 3.7).

6. Construct the resulting arrangement graph. Note that the graph is directed and acyclic, except for the trivial subgraphs around S and T (see Section 3.3).

7. Use topological sort to find the shortest path.

Note that we have three approximation steps applied one after another (exponential subarrangement and the two critical triangles) for a factor of $(1+\varepsilon')$ each. The total approximation factor is $(1+\varepsilon')^3 = (1+\varepsilon)$ as desired.

3.2 Simple arrangements

An arrangement of lines $\mathscr{A}$ containing points S and T is *simple* if

1. Every exterior line is a boundary line—no lines are outside $\mathrm{Hull}(S,T)$.

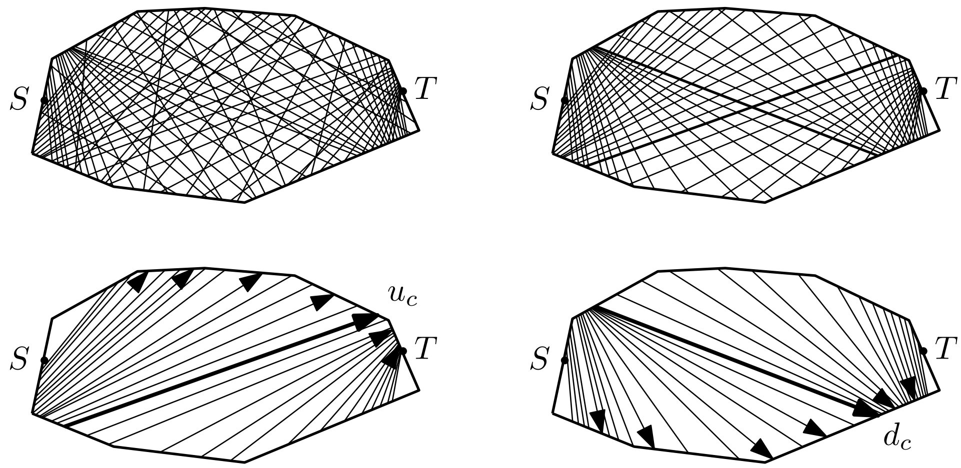


Figure 3.1: Top row: an arrangement and its simple subarrangement. Bottom row: up-lines and down-lines of the simple subarrangement.

2. No two up-lines (down-lines) intersect within $\mathrm{Hull}(S,T)$.

   We define a total order on all up-lines (down-lines): $a <_* b$ iff any path from S to T that does not leave $\mathrm{Hull}(S,T)$ crosses a before b. We label the up-lines $\{u_1, u_2, \ldots, u_n\}$ where $i < j \iff u_i <_* u_j$. Similarly for the down-lines $\{d_1, d_2, \ldots, d_m\}$.

3. There's a unimodal ordering of steepness for all up-lines (down-lines). That is, $u_1 > u_2 > \ldots > u_{c-1} > u_c < u_{c+1} < \ldots < u_n$ (similarly for down-lines). We call the shallowest up-line u_c the *critical up-line* (similarly for the *critical down-line*).

   Note: the c in u_c and d_c is used for convenience and not meant as a specific index (which would imply that u_c and d_c are at exactly the same position in their respective lists).

Any arrangement $\mathscr{A}$ has a *simple subarrangement* constructed by the following procedure:

1. Discard the lines outside of $\mathrm{Hull}(S,T)$.

2. Order the up-lines by the intersections along the lower boundary to obtain

the list $\{u_1, u_2, \ldots, u_n\}$. We will construct the new set of up-lines $\mathscr{U}$ starting with $\{u_c\}$:

For every u_i in the sequence $(u_{c-1}, u_{c-2}, \ldots, u_1)$:

Discard from $\mathscr{U}$ any lines $\{u_k | u_i < u_k\}$; then add u_i to $\mathscr{U}$.

For every u_i in the sequence $(u_{c+1}, u_{c+2}, \ldots, u_n)$:

Discard from $\mathscr{U}$ any lines $\{u_k | k > c; u_i < u_k\}$; then add u_i to $\mathscr{U}$.

3. Construct the new set of down-lines in a similar fashion.

Note that we can construct the simple subarrangement in time $O(n \log n)$ which is dominated by the time required to construct $\mathrm{Hull}(S,T)$ and to order the intersections with the boundary; if we use a stack to construct $\mathscr{U}$, we would only ever discard lines from the top of the stack for a total of $O(n)$.

The fact that it's easy to construct and the following result allow us to look exclusively at simple arrangements:

Theorem 3.2.1. *If $\mathscr{A}_s$ is a simple subarrangement of $\mathscr{A}$ and p is a shortest path in $\mathscr{A}_s$, then p is a shortest path in $\mathscr{A}$.*

Proof. We show that no lines removed in the construction of the subarrangement would have been used by the shortest path.

The lines outside of $\mathrm{Hull}(S,T)$ are not used by the shortest path.

Consider a line u_k discarded from $\mathscr{U}$ during the construction procedure because $u_i < u_k$ for some $i < k < c$. If $u_i \cap u_k \in \mathrm{Hull}(S,T)$ then the shortest path would not travel on u_k by Lemma 2.6.1. Otherwise the shortest path would not travel on u_k by Lemma 2.5.1, as any segment of travel on u_k would be located between two less steep lines: u_i and u_c.

The proof for the case $c < k < i$ and for the down-lines is similar. $\square$

3.3 Cycles in simple arrangements

It's possible for the directed graph induced by a simple arrangement to contain cycles in it (see Figure 3.2) if an up-line and a down-line cross at an angle larger

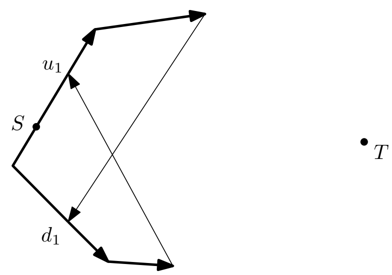


Figure 3.2: A possible cycle between two cross lines and the boundary.

than π. Note that the shortest path would not use the intersection between these two lines by Observation 2.1.1.

Due to the unimodal ordering of line angles, these "retrograde" regions are contiguous, are located adjacent to S or T and are easy to compute (see Figure 3.3). Since we can ignore the interior intersections for the purpose of finding the shortest path, the subgraph contains $O(n)$ vertices (all of them on the boundary), and the shortest path to each vertex is found in $O(n \log n)$.

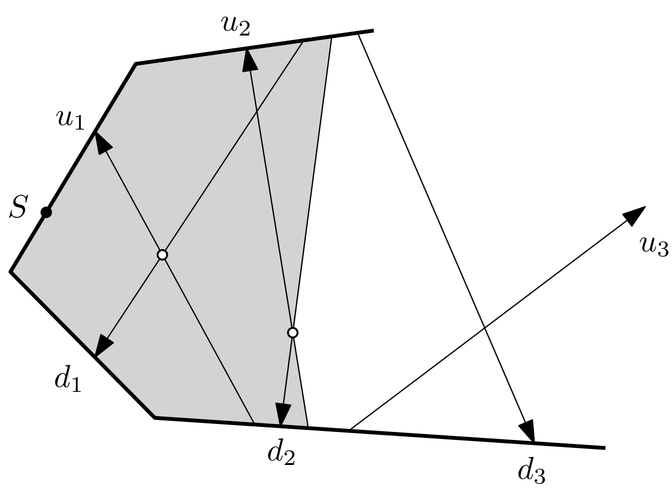


Figure 3.3: The retrograde region around S and the unused cross line intersections.

3.4 Arrangement quadrants

It's useful to partition Hull(S,T) into regions formed by the lines u_c and d_c[1] (see Figure 3.4). The region below d_c and above u_c is called the *S quadrant*; the region above d_c and below u_c is called the *T quadrant*. The region above both lines is called the *upper quadrant*, while the region below both lines is called the *lower quadrant*.

It is easy to compute the shortest path in the upper and lower quadrants: we show that the shortest path does not turn from one cross line to another.

The S and T quadrants require additional approximation steps, detailed below.

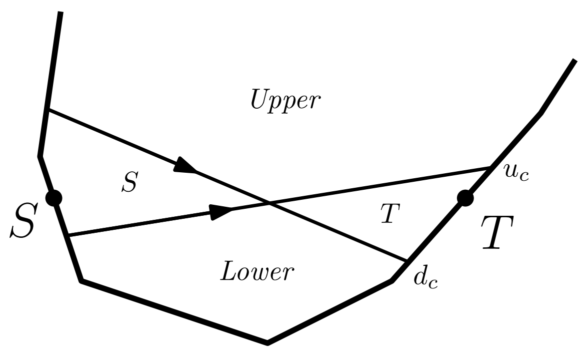


Figure 3.4: Arrangement quadrants.

3.5 Upper and lower quadrants

Theorem 3.5.1. *A shortest path p in a simple arrangement does not turn from one cross line to another after leaving d_c and before arriving to u_c (that is, in the upper quadrant) or after leaving u_c and before arriving to d_c (that is, in the lower quadrant).*

Proof. Suppose, without loss of generality, that p leaves d_c, turns at a point Q from a cross line b to a cross line c, and then arrives at u_c. Let a be the line on which p arrives at b and d be the line on which p leaves c. Let $A = a \cap b$, $B = c \cap d$, and $P = a \cap d$.

[1] u_c and d_c might not intersect within Hull(S,T), in which case the upper or lower quadrant could be empty.

Case 1. b is an up-line and c is a down-line (see Figure 3.5).

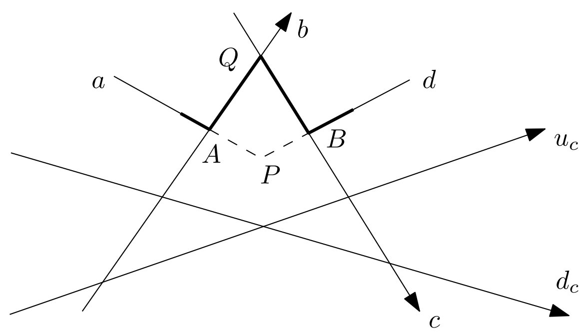


Figure 3.5: Case 1. Eliminating travel on b and c results in a shorter path.

It must be that $a < c$ since if $a > c$ then also $a > d_c$ and by Lemma 2.5.1 the shortest path would not travel on a.

Similarly, it must be that $d < b$.

Consider the quadrilateral $AQBP$. Since $a < c$ and $d < b$ we have $d(A,P,B) < d(A,Q,B)$, so we obtained a shorter path, a contradiction.

Case 2. b is a down-line and c is an up-line. Let $C = d \cap u_c$ and $D = b \cap u_c$. By Case 1, we must conclude that d is a boundary line—otherwise the path would make a cross line to cross line turn at B. Let U be the point at which the path arrives at u_c; it must be that $U \in \overline{DC}$.

Consider the quadrilateral $QBCD$. If $d > b$ then $d(Q,D,U) < d(Q,B,U) \leq |p_{QU}|$ where p_{QU} is the actual subpath from Q to U, so we must conclude that $d < b$ (see Figure 3.6).

By a similar argument we conclude that $a < c$.

Now as in the first case, considering the quadrilateral $AQBP$ we see that $d(A,P,B) < d(A,Q,B)$ and we have a contradiction (see Figure 3.7).

□

3.6 S and T quadrants

3.6.1 Exponential subarrangements

Any simple arrangement $\mathscr{A}$ has an *exponential subarrangement* $\mathscr{A}_\varepsilon$ defined below.

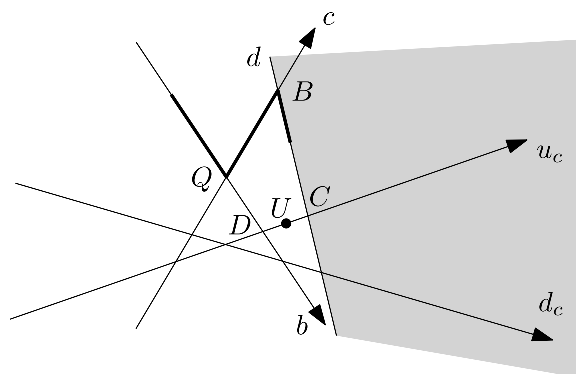


Figure 3.6: Case 2. If $d > b$, $\overline{QD}$ and $\overline{DU}$ give a shorter path.

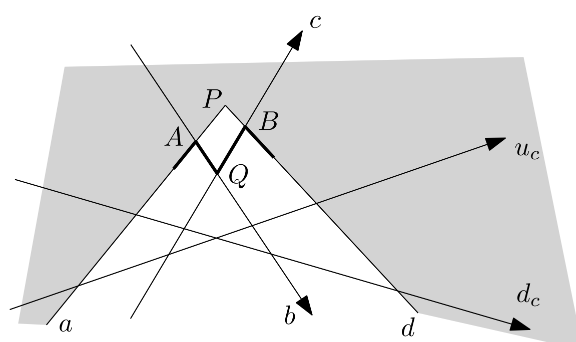


Figure 3.7: Case 2. If $d < b$ and $a < c$, $\overline{AP}$ and $\overline{PB}$ give a shorter path.

Let $\mathscr{A}$ be a simple arrangement. As before, let the up-lines be labeled as $\{u_1, u_2, \ldots, u_n\}$ where $i < j \iff u_i <_* u_j$ with u_c the critical up-line. Similarly for the down-lines and d_c.

Let σ_i (τ_i) be the angle that the directed line u_i makes with the line through S (line through T) considered as a directed line with a positive y-coordinate travel. If u_i makes a negative angle with the line through S (line through T), then the shortest path would not use it by Lemma 2.5.1, so we assume that $\sigma_i, \tau_i \in (0, \pi)$. Note that for steep up-lines, σ_i and τ_i are small. We will use the notation $\angle u_i = \sigma_i$.

$\mathscr{A}_\varepsilon$ is constructed by the following procedure:

1. Construct the new set of up-lines $\mathscr{U}_\varepsilon$ starting with $\{u_c\}$:

   For every u_i in the sequence $(u_{c-1}, u_{c-2}, \ldots, u_1)$:

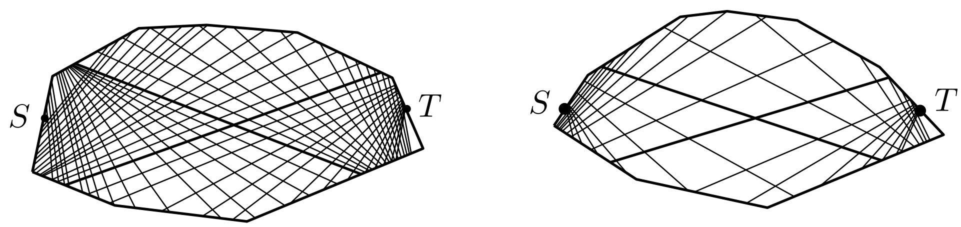


Figure 3.8: A simple arrangement and its exponential subarrangement. Note that the removed cross lines can still be used as *bridge segments* to jump from boundary to boundary, but not to other cross lines.

Let u_j be the last up-line added to $\mathscr{U}_\varepsilon$. Add u_i to $\mathscr{U}_\varepsilon$ if $i = 1$ or $\sigma_j/\sigma_{i-1} > 1+\varepsilon/2$.

For every u_i in the sequence $(u_{c+1}, u_{c+2}, \ldots, u_n)$:

Let u_j be the last up-line added to $\mathscr{U}_\varepsilon$. Add u_i to $\mathscr{U}_\varepsilon$ if $i = n$ or $\tau_j/\tau_{i+1} > 1+\varepsilon/2$.

2. Construct the new set of down-lines in a similar fashion.

3. Augment $\mathscr{A}_\varepsilon$ with a set of *bridge segments*. For every up-line and down-line in $\mathscr{A}$ not already included, we include the line segment connecting the lower boundary to the upper. The shortest path is not allowed to turn from a cross line to a bridge segment or vice versa, or from one bridge segment to another.

Another way to view an exponential subarrangement is as keeping all the lines in the arrangement and removing some cross line to cross line intersections (we call these cross lines bridge segments).

It's easy to see that we can construct $\mathscr{A}_\varepsilon$ in time $O(n \log n)$, the time needed to order the lines by σ_i and τ_i. The intention is to get rid of lines that are close to each other and have similar angles. We show that this does not significantly increase the length of the shortest path.

3.6.2 Line sequences

It will help for the discussion below to develop a language when we talk about the sequence of line types traversed by a shortest path.

We will use the symbols u, d, U, D to refer to an arbitrary up-, down-, upper boundary, and lower boundary line respectively. If we need to label a specific line, we will use u_1, d_2, etc.

A sequence of symbols is *valid* if it follows the following rules:

1. u or U can be immediately followed by d or U

2. d or D can be immediately followed by u or D

For example, udDDuUUdu is a valid sequence, while uddUD is not.

We can identify any shortest path p with a string of the above symbols corresponding to each line that p travels on in order from S to T. By Observation 2.1.1, any shortest path will map to a valid sequence.

A valid sequence is *exact* if no u is immediately followed by d and vice versa. For example, dDDuUUd is an exact sequence, while dDDudDu is valid but not exact. Paths that map to exact sequences travel along the boundary or cross $\mathrm{Hull}(S,T)$ using bridge segments without turning from one cross line to another. Any such path would be completely preserved in an exponential subarrangement, thus the name.

An *upper subsequence* of a given sequence is the subsequence of all u and U elements with their relative order preserved. Similarly, a *lower subsequence* is the subsequence of all d and D elements. A *reordering* is the sequence obtained by concatenating the upper subsequence with the lower subsequence. For example, a sequence $U_1d_2D_3D_4u_5U_6U_7d_8$ has an upper subsequence $U_1u_5U_6U_7$ and a lower subsequence $d_2D_3D_4d_8$; its reordering is the sequence $U_1u_5U_6U_7d_2D_3D_4d_8$.

We will talk about the *reordering of a path* in the sense of taking the individual line segments that the path travels on and reordering them into another contiguous poly-line (by translating the line segments while preserving the angles) according to the path's sequence reordering. Note that a path reordering creates a poly-line from the original path's starting to its ending point of the same length as the original path.

Finally we have an observation about the line angles in the sequence which mirrors the unimodal angle ordering in simple arrangements.

Observation 3.6.1. *Let p be a shortest path in a simple arrangement. If p travels on an up-line u_j before reaching u_c, then for all $i < j$ in the upper subsequence of p, $l_i > u_j$ (l_i could be an upper boundary line or an up-line).*

Similarly for the down-lines.

Proof. If there's a line $l_i < u_j$ where $i < j$ then by Lemma 2.5.1 the path would not have travelled on u_j (any travel segment is between the shallower lines l_i and u_c), so it must be that $l_i > u_j$. $\square$

3.6.3 Error-wedge Lemma

We will make use of the following lemma, which allows us to approximate shortest path segments with travel on two lines:

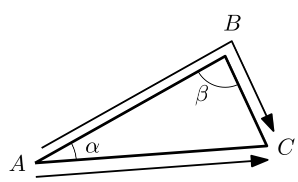


Figure 3.9: Error-wedge Lemma.

Lemma 3.6.2 (Error-wedge Lemma)**.** *Let $\triangle ABC$ be a triangle with $\alpha = \angle BAC$ and $\beta = \angle ABC$. If $\alpha/\beta < \varepsilon/2$ then $d(A,B,C) < (1+\varepsilon)d(A,C)$.*

Proof. First, we need the fact that $f(\beta) = \frac{\beta(\cos(\beta)+1)}{\sin(\beta)} < 2$ for $\beta \in (0,\pi)$. We omit full details, but this is easily shown since $\lim_{\beta\to 0^+} f(\beta) = 2$, $\lim_{\beta\to\pi^-} f(\beta) = 0$, and $\frac{d}{d\beta} f(\beta) < 0$ for $0 < \beta < \pi$.

Using further the fact that $d(A,B) = d(A,C)\cos(\alpha) + d(B,C)\cos(\beta)$, as well as the law of sines $d(A,C)\sin(\alpha) = d(B,C)\sin(\beta)$, we have

$$
\begin{aligned}
\frac{d(A,B,C)}{d(A,C)} &= \\
\frac{d(A,C)\cos(\alpha)+d(B,C)\cos(\beta)+d(B,C)}{d(A,C)} &= \\
\frac{d(A,C)\cos(\alpha)+d(B,C)(\cos(\beta)+1)}{d(A,C)} &= \\
\frac{d(A,C)\cos(\alpha)+d(A,C)\frac{\sin(\alpha)}{\sin(\beta)}(\cos(\beta)+1)}{d(A,C)} &= \\
\cos(\alpha)+\frac{\sin(\alpha)}{\sin(\beta)}(\cos(\beta)+1) &= \\
\cos(\alpha)+\frac{\sin(\alpha)}{\beta}\left(\frac{\beta(\cos(\beta)+1)}{\sin(\beta)}\right) &< \\
1+\frac{\alpha}{\beta}\left(\frac{\beta(\cos(\beta)+1)}{\sin(\beta)}\right) &< \\
1+\frac{\alpha}{\beta}2 &< \\
1+\varepsilon &
\end{aligned}
$$

□

3.6.4 $(1+\varepsilon)$ approximation

Lemma 3.6.3. *Given a simple arrangement $\mathscr{A}$, let $\mathscr{A}_\varepsilon$ be its exponential subarrangement, and P a point on an up-line, down-line, or a boundary line in the S quadrant of $\mathscr{A}_\varepsilon$. If p is a shortest S–P path in $\mathscr{A}$ and q is a shortest S–P path in $\mathscr{A}_\varepsilon$, then $|q| < (1+\varepsilon)|p|$.*

Proof. We prove this by induction on (n_u, n_d)—the number of up-lines and down-lines in $\mathscr{A}_\varepsilon$ respectively that any shortest S–P path will intersect.

Without loss of generality, let P be a point on a down-line or an upper boundary line d_k in A_ε.

Let u_m be the last up-line in $\mathscr{A}_\varepsilon$ crossed by p (we take the boundary line through S to be u_m if no such up-line exists). Let u_{m+1} be the next up-line in $\mathscr{A}_\varepsilon$, which must exist and $u_m > u_{m+1} \geq u_c$ since P is in the S quadrant and p does not cross u_c.

Let U be the point where p leaves u_m and let p' be the U–P subpath of p. By induction, the S–U subpath is well-approximated, so it's sufficient to show that the U–P subpath is also well-approximated. We assume that p' travels on at least two lines—otherwise the single line must be u_m or an upper boundary line by Observation 2.1.1, and we apply induction by removing d_k.

Consider the maximal contiguous exact suffix s of p' ending at P, excluding any initial u or d in the suffix. By construction, every line in s is in $\mathscr{A}_\varepsilon$ since we keep all boundary lines and bridge segments.

Let Q be the point where p' enters s. If s is nonempty, Q must lie on a boundary line and by the inductive assumption we can well-approximate the shortest S–Q path, which does not intersect d_k and thus intersects fewer lines than the shortest S–P path. So assume that s is empty—that is, $Q = P$.

Note that the assumption of travel on at least two lines and the fact that s is empty implies that p' ends with either ud or du.

If $\mathscr{A}$ has no up-lines between u_m and u_{m+1} then u in the sequence must be u_m, and the only possibility is that p' travels on u_m followed by d_k. We remove d_k and proceed by induction.

Otherwise, there's at least one up-line between u_m and u_{m+1} that we skipped during the construction of $\mathscr{A}_\varepsilon$, and we must have $\sigma_{m+1}/\sigma_m < 1 + \varepsilon/2$.

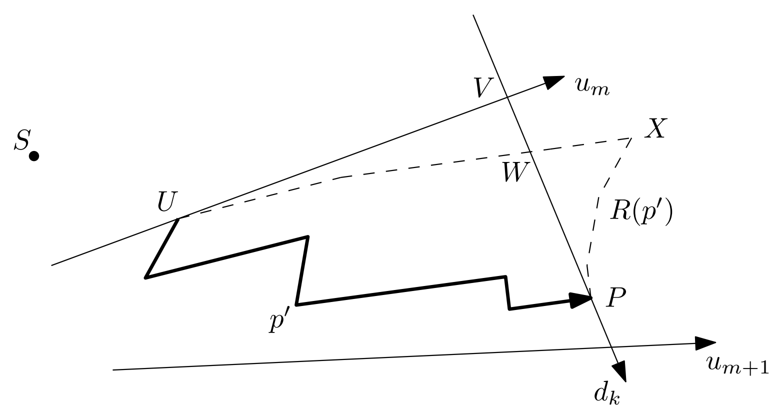


Figure 3.10: $d(U,V,P) < (1+\varepsilon)d(U,W,P) \leq (1+\varepsilon)|p'|$.

Consider the reordering of p', and let X be the point at the end of the upper subsequence in the reordering. We would like to show that (1) d_k separates U from

X, and that (2) all lines in the upper subsequence of p are more steep than u_{m+1}. These two facts will allow the application of Lemma 3.6.2.

Since p' ends with either ud or du, its upper subsequence ends with u, and its lower subsequence ends with d. By Observation 3.6.1, we immediately have (2); and, since all lines in its lower subsequence are steeper than d_k, we have (1) by convexity of the reordering.

Let $V = u_m \cap d_k$, and let W be the point where the reordering of p' first crosses d_k. Since $d(U,W,P) \leq |p'|$ and $\overline{WP} \in d_k$, it's sufficient to show that travel on $\overline{UV} \in u_m$ followed by $\overline{VW} \in d_k$ well-approximates p'.

Note that by (1) W must be within the upper subsequence of the reordering. By (2) we have $\angle VUW < \sigma_{m+1} - \sigma_m$, and we have

$$\frac{\angle VUW}{\angle UVW} < \frac{\sigma_{m+1} - \sigma_m}{\sigma_m + \delta_k} < \frac{\sigma_{m+1} - \sigma_m}{\sigma_m} = \frac{\sigma_{m+1}}{\sigma_m} - 1 < \frac{\varepsilon}{2}$$

Finally by Lemma 3.6.2 we have $d(U,V,W) + d(W,P) < (1+\varepsilon)d(U,W) + d(W,P) < (1+\varepsilon)d(U,W,P) \leq (1+\varepsilon)|p'|$. $\square$

3.7 Approximation in the critical triangle

Let i and j be the largest indices so that $\angle u_i < \min\{\pi/4, \varepsilon/(12+6\varepsilon)\angle u_c\}$ and $\angle d_j < \min\{\pi/4, \varepsilon/(12+6\varepsilon)\angle d_c\}$. We call the triangle defined by u_i, d_j, and s the *critical triangle* and show that there's a particularly simple approximation for the part of the shortest path that traverses it.

Note that to get the full approximation we need to apply the lemma below twice (for the critical triangle next to S and T) for a total approximation factor of $(1+\varepsilon)^2$. Taking this together with another factor of $(1+\varepsilon)$ from the exponential subarrangement, we would use $\varepsilon' = \sqrt[3]{1+\varepsilon} - 1$ in the complete algorithm instead.

Lemma 3.7.1. *Given a unit circle $\mathscr{U}$, let a, b, and c be three lines tangent to it so that $Q = a \cap b$ is not in the same half-plane defined by c as $\mathscr{U}$. Let $P = a \cap c$ and $R = b \cap c$. Let $\alpha = \angle PQR \in (0, \pi)$.*

Then $\partial(\triangle PQR) > (\pi - \alpha)/2$. *($\partial(x)$ is the perimeter of x)*

If $\alpha > \pi/2$ then also $\partial(\triangle PQR) < 3(\pi - \alpha)$.

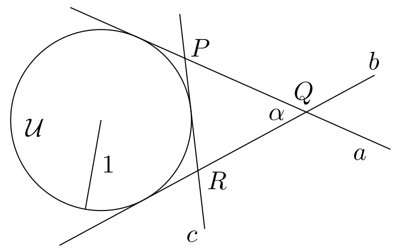


Figure 3.11: $(\pi - \alpha)/2 < \partial(\triangle PQR) < 3(\pi - \alpha)$. The upper bound holds only if $\alpha > \pi/2$.

Proof. For the lower bound, using the Taylor series expansion, we have for $0 < x < \pi/2$:

$$\sin(x) > x - x^3/6$$

Let $\gamma = (\pi - \alpha)/2 < \pi/2$.

$$\partial(\triangle PQR) = 2\tan(\gamma) = 2\frac{\sin(\gamma)}{\cos(\gamma)} >$$

$$2\frac{\gamma - \gamma^3/6}{1} > \gamma = (\pi - \alpha)/2$$

For the upper bound, using the Taylor series expansions again, we have for $0 < x < \pi/4$

$$\sin(x) < x - x^3/6 + x^5/5! < x + x^5/5!,$$

and

$$\cos(x) > 1 - x^2/2 > 1 - (\pi/4)^2/2 > 17/25.$$

Let $\gamma = (\pi - \alpha)/2 < \pi/4$.

$$\partial(\triangle PQR) = 2\tan(\gamma) = 2\frac{\sin(\gamma)}{\cos(\gamma)} <$$

$$2\frac{\gamma + \gamma^5/5!}{17/25} < 2\frac{2\gamma}{17/25} = 100/17\gamma < 6\gamma = 3(\pi - \alpha)$$

$\square$

Corollary 3.7.1.1. *Given a circle $\mathscr{C}$, let a, b, and c be three lines tangent to it so* *that $X = a \cap b$ is not in the same half-plane defined by c as $\mathscr{C}$. Let $Y = a \cap c$ and $Z = b \cap c$.*

Let further d and e be two lines also tangent to $\mathscr{C}$ so that $Q = d \cap e$ is not in the same half-plane defined by c as $\mathscr{C}$.

If $\angle RQP > \pi/2$ and $(\pi - \angle RQP) < \varepsilon(\pi - \angle YXZ)$, then $\partial(\triangle PQR) < 6\varepsilon\partial(\triangle XYZ)$.

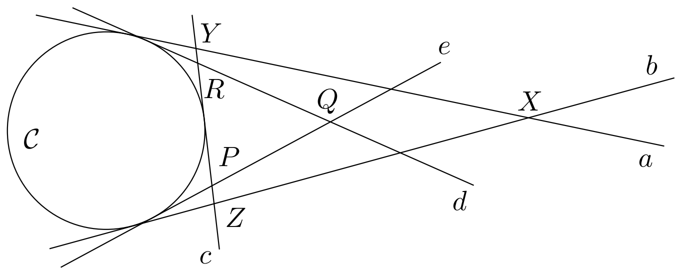


Figure 3.12: $\partial(\triangle PQR) < 6\varepsilon\partial(\triangle XYZ)$.

Proof. Since we're only concerned about ratios, we can assume that $\mathscr{C}$ is a unit circle without loss of generality.

By Lemma 3.7.1, $\partial(\triangle PQR) < 3(\pi - \angle RQP)$ and $\partial(\triangle XYZ) > (\pi - \angle YXZ)/2$. We have

$$\partial(\triangle PQR) < 3(\pi - \angle RQP) <$$
$$3\varepsilon(\pi - \angle YXZ) = 6\varepsilon(\pi - \angle YXZ)/2 <$$
$$6\varepsilon\partial(\triangle XYZ)$$

□

Lemma 3.7.2 (Tangent-bound Lemma)**.** *Let E be a simple arrangement. If p is a shortest path in E and q is a shortest path in E but with the restriction that no turns are allowed between two cross lines within the critical triangle, then $|q| < (1+\varepsilon)|p|$.*

Proof. If p doesn't turn from a cross line to a cross line within the critical triangle then we're done.

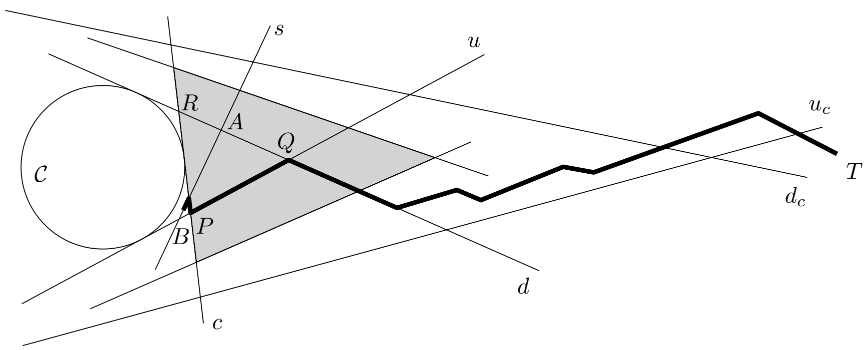


Figure 3.13: p last turns at Q in the critical triangle.

Otherwise, let Q be the point at the last such turn when traversing p from S towards T. Without loss of generality, let's assume that p arrives at Q on an up-line u and leaves Q on a down-line d, and also that p arrives to u on a line c (which could be a down-line or a lower boundary line by Observation 2.1.1); let $P = c \cap u$.

We would like to apply the Lemma 2.4.1 to the shortest path from S to T travelling on the lines c, u (which is a cross line by assumption and thus separates S from T), and d. It must be that $c > d$ since otherwise $c < d > d_c$ and the shortest path would not use d by Lemma 2.5.1. So this fixes the circle in the Circle Lemma definition to be the unique circle $\mathscr{C}$ tangent to c, u, and d that lies in the half-plane defined by c that does not contain Q. Let also $R = c \cap d$.

We have that s (the line through S) does not cross $\mathscr{C}$.

So it must be that s intersects d within the closed segment $\overline{RQ}$. Let $A = s \cap d \in \overline{RQ}$ and $B = s \cap u$.

We also get that no lines crossed by the shortest path after leaving Q cross $\mathscr{C}$—in particular, u_c and d_c do not cross $\mathscr{C}$.

Let the approximate path q be the path obtained from p by replacing the existing subpath from S to Q with the two segments $\overline{SA}$ and $\overline{AQ}$. Note that q does not turn from a cross line to a cross line within the critical triangle. We claim that q provides the required approximation.

Let u_c' be a line parallel to u_c and tangent to $\mathscr{C}$ (we choose the tangent that puts u_c and $\mathscr{C}$ in the different half-planes defined by u_c'). We define d_c' in a similar way. Let $V = u \cap d_c'$, $X = d_c' \cap u_c'$, $Y = d_c' \cap c$, and $Z = u_c' \cap c$ (see Figure 3.14).

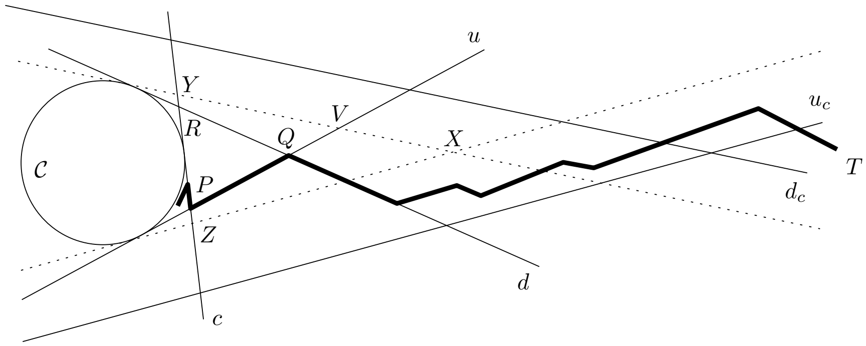


Figure 3.14: Parallel tangents.

Let $\delta = |q| - |p|$. It's sufficient to show that $\delta < \varepsilon|p|$, which we prove in a few steps:

1. $\delta \leq \partial(\triangle PQR)$

   This follows from the facts that $\delta \leq \partial(\triangle AQB)$ and $\partial(\triangle AQB) \leq \partial(\triangle PQR)$ (the perimeters are equal if s is tangent to $\mathscr{C}$).

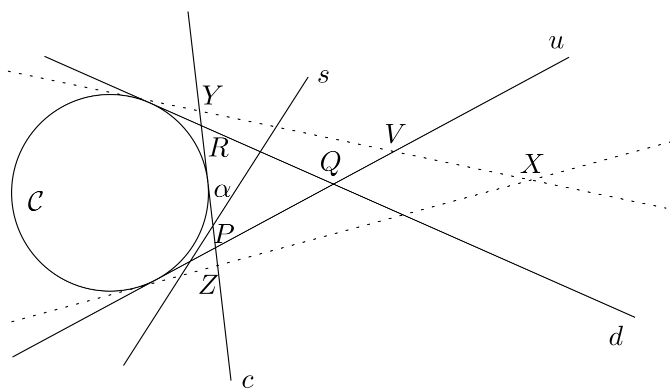


Figure 3.15: $\partial(\triangle PQR) < \varepsilon d(Q,V,X)$.

2. $\partial(\triangle PQR) < \varepsilon d(Q,V,X)$

   Proof. Since $\angle RPQ = \angle u + \alpha$ and $\angle PRQ = \angle d - \alpha$, where α is the angle between c and s, we have $\pi - \angle RQP = \angle PRQ + \angle RPQ = \angle u + \angle d$ (see Figure 3.15).

Similarly, $\pi - \angle YXZ = \angle u_c + \angle d_c$.

Since u and d are in the critical triangle, we have $\angle u + \angle d < \varepsilon/(12+6\varepsilon)(\angle u_c + \angle u_d)$ and $\angle u + \angle d < \pi/2$. We have

$$\pi - \angle RQP < \frac{\varepsilon}{12+6\varepsilon}(\pi - \angle YXZ) \qquad \angle RQP > \pi/2$$

$$\partial(\triangle PQR) < \frac{\varepsilon}{2+\varepsilon}\partial(\triangle XYZ) \quad \text{(By Corollary 3.7.1.1)}$$

$$\partial(\triangle XYZ) - \partial(\triangle PQR) = 2d(Q,V,X)$$

$$\frac{2+\varepsilon}{\varepsilon}\partial(\triangle PQR) - \partial(\triangle PQR) < 2d(Q,V,X)$$

$$\frac{1}{\varepsilon}\partial(\triangle PQR) < d(Q,V,X)$$

□

3. $d(Q,V,X) \leq |p|$

This follows immediately from the following intuitively obvious claim:

Given a circle $\mathscr{C}$, let b_1 and b_2 be the tangents of $\mathscr{C}$ through B. Let A be a point in the region bounded by b_1, b_2, and $\mathscr{C}$, and let a_1 and a_2 be the tangents of $\mathscr{C}$ through A.

Let $ACBD$ be the quadrilateral defined by the intersections of the four tangents. Note that by Lemma 2.3.4, $d(A,C,B) = d(A,D,B)$.

If p is a shortest A–B path that doesn't travel on any lines that cross $\mathscr{C}$ then $|p| \geq d(A,C,B) = d(A,D,B)$ (see Figure 3.16).

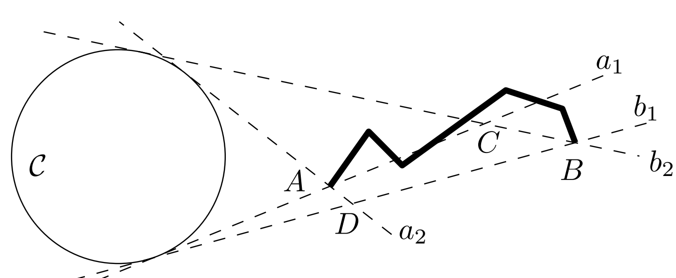


Figure 3.16: Any shortest A–B path that respects $\mathscr{C}$ must be at least as long as $d(A,C,B) = d(A,D,B)$.

Proof. If $A \in b_1$ or $A \in b_2$, the result follows since $d(A,C,B) = d(A,B)$, and any shortest path is at least as long as $d(A,B)$.

If p only travels on a single segment $\overline{AB}$, then it must be that $A \in b_1$ or $A \in b_2$ (otherwise $\overleftrightarrow{AB}$ would cross $\mathscr{C}$), and the result follows as above.

Otherwise, let $\overline{AX}$ be the first segment of p. If this segment intersects b_1 or b_2 we take X to be this point of intersection instead.

Let x_1 and x_2 be the tangents of $\mathscr{C}$ through X, and let $XYBZ$ be the quadrilateral defined by the four tangents b_1, b_2, x_1, and x_2. By Lemma 2.3.4, $d(X,Y,B) = d(X,Z,B)$ (see Figure 3.17).

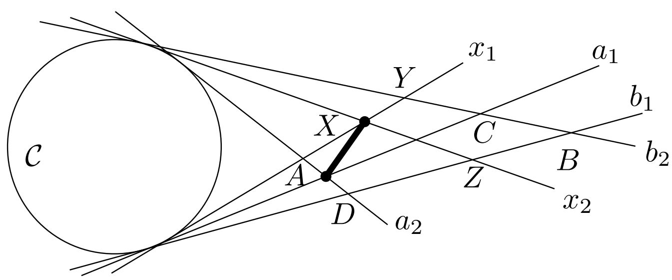


Figure 3.17: Induction step. $d(A,X,Y,B) \geq d(A,C,B)$.

By induction on the number of line segments in p, $|p| = d(A,X) + |p_{XB}| \geq d(A,X) + d(X,Y,B)$ (note that the induction must terminate in the case above when p crosses b_1 or b_2).

Since $\overleftrightarrow{AX}$ does not cross $\mathscr{C}$, then either $C \in \overline{YB}$ or $D \in \overline{ZB}$. Since $d(X,Y,B) = d(X,Z,B)$, we can assume without loss of generality that $C \in \overline{YB}$.

We have $d(A,X) + d(X,Y,B) = d(A,X) + d(X,Y,C,B) = d(A,X,Y,C) + d(C,B) \geq d(A,C,B)$ by the triangle inequality. □

Finally, putting the steps together we get $\delta < \varepsilon|p|$ and the shortest path is well-approximated by q. □

3.8 Running time

Ordering the lines by angle and choosing the lines that are part of the exponential subarrangement can be done in $O(n \log n)$ time.

The graph induced by the arrangement will contain $O(n)$ edges, but we do not need to consider turns between cross lines within the critical triangles or the upper and lower regions so we can save on the number of vertices.

We will use ε' where $(1+\varepsilon')^3 = (1+\varepsilon)$ for the exponential subarrangement and the critical triangle definitions to simplify the discussion.

Let's first calculate the number of up-lines inside and outside the critical triangle in the S quadrant. Suppose that the total number of up-lines in the S quadrant is k and the critical triangle includes the lines $u_1, \ldots, u_i$ with $\angle u_i < x < \angle u_{i+1}$, where $x = \min\{\pi/4, (\varepsilon'/(12+6\varepsilon'))\angle u_c\}$.

By the exponential subarrangement construction, $\angle u_j / \angle u_{j-2} \geq 1 + \varepsilon'/2$ for all $3 \leq j \leq c$. We have

$$\frac{\angle u_j}{\angle u_{j-2m}} = \frac{\angle u_j}{\angle u_{j-2}} \times \frac{\angle u_{j-2}}{\angle u_{j-4}} \times \cdots \times \frac{\angle u_{j-2m+2}}{\angle u_{j-2m}} \geq (1+\varepsilon'/2)^m$$

$$\angle u_i < (\varepsilon'/(12+6\varepsilon'))\angle u_c$$

$$\frac{\angle u_c}{\angle u_i} > 12/\varepsilon' + 6$$

$$\frac{\angle u_c}{\angle u_i} \geq \frac{\angle u_c}{\angle u_{c-2m}} \geq (1+\varepsilon'/2)^m > 12/\varepsilon' + 6$$

$$m > \frac{\log(12/\varepsilon' + 6)}{\log(1+\varepsilon'/2)}$$

In other words, if m is at least this large, the remaining $k-m$ lines are in the critical triangle. In total, we have $m^2 + 2m(k-m) = 2mk - m^2$ vertices in the graph in the S quadrant.

The T quadrant is similar, and the top and bottom quadrants do not contribute any vertices. Asymptotically, since $k = O(n)$, we get $|V| = O(mn)$.

1. $m = O(1/e' \log(1/e'))$

   Let $\phi' = 1/\varepsilon'$, so that $\log(12/\varepsilon' + 6)/\log(1+\varepsilon'/2) = \log(12\phi' + 6)/\log(1+$

$1/(2\phi'))$. If $g(\phi') = O(f(\phi'))$ then there are C, ϕ_0' so that for all $\phi' > \phi_o'$:

$$\begin{aligned}
\frac{\log(12\phi'+6)}{\log(1+1/(2\phi'))} &< Cf(\phi') \\
\log(12\phi'+6) &< Cf(\phi')\log(1+1/(2\phi')) \\
12\phi'+6 &< e^{Cf(\phi')\log(1+1/(2\phi'))} && \text{exponentiate} \\
12\phi'+6 &< \phi'^{C\phi'\log(1+1/(2\phi'))} && \text{take } f(x) = x\log x
\end{aligned}$$

Using $\log(1+x) = x - x^2/2 + x^3/3 - \ldots$:

$$\begin{aligned}
12\phi'+6 &< \phi'^{C\phi'(1/(2\phi') - (1/(2\phi'))^2/2 + (1/(2\phi'))^3/3 - \ldots)} \\
12\phi'+6 &< \phi'^{C(1/2 - 1/(2\phi') + (1/(3\cdot 8\phi'^2) - \ldots)}
\end{aligned}$$

Which holds for ϕ' and C sufficiently large to make the exponent greater than one, so $\log(12\phi'+6)/\log(1+1/(2\phi')) = O(\phi'\log\phi')$.

2. $1/\varepsilon' = O(1/\sqrt[3]{\varepsilon})$. Since $\varepsilon' = \sqrt[3]{1+\varepsilon} - 1$, we have

$$\begin{aligned}
\phi' &= \frac{1}{\sqrt[3]{1+1/\phi} - 1} \\
\phi' &= \frac{\sqrt[3]{\phi}}{\sqrt[3]{\phi+1} - \sqrt[3]{\phi}} && \text{multiplying through by } \sqrt[3]{\phi} \\
\phi' &= O(\sqrt[3]{\phi})
\end{aligned}$$

3. It follows that $m = O(\varepsilon^{-3}\log(\varepsilon^{-1}))$.

Since the graph is directed and acyclic[2], we can simply use the $O(n)$ topological sort to find the shortest path, and the overall time complexity is

$$\begin{aligned}
O(n\log n) + O(|V|) = O(n\log n + mn) = \\
O\left(n\log n + \varepsilon^{-3}\log(\varepsilon^{-1})n\right)
\end{aligned}$$

□

[2]except for the trivial subgraph, where the shortest path is also found in $O(n\log n)$—see Section 3.3

Chapter 4

Future work

The problem and some of our proofs suggest a number of possible directions to explore.

The Circle Lemma seems to be fundamental in understanding the problem. We believe that other results, like the Phantom Lemma and the Two-lines Lemma could be restated as corollaries of the Circle Lemma. The Circle Lemma reduces the shortest path problem to a potentially simpler combinatorial problem of finding if certain lines intersect certain circles.

The fact that the best known solution is $O(n^2)$ suggests a possibility that this problem falls into a class of "3SUM-hard" problems [7][14]. It is also not known if any of these problems can be solved in sub-$O(n^2)$ time.

Other avenues to explore is finding a sub-$O(n^2)$ exact solution on a restricted problem (for example, with bounds on minimum differences between line angles or distances between intersections in the arrangement), or a sub-$O(n \log n)$ approximation.

Bibliography

[1] P. Bose, W. S. Evans, D. G. Kirkpatrick, M. McAllister, and J. Snoeyink. Approximating shortest paths in arrangements of lines. In *CCCG*, volume 96, pages 143–148. Citeseer, 1996. → page 2

[2] D. Z. Chen, S. Luan, and J. Xu. Topological peeling and implementation. In *International Symposium on Algorithms and Computation*, pages 454–466. Springer, 2001. → page 2

[3] D. Z. Chen, O. Daescu, X. S. Hu, and J. Xu. Finding an optimal path without growing the tree. *Journal of Algorithms*, 49(1):13–41, 2003. → page 2

[4] C. Davis. The short-cut problem. *The American Mathematical Monthly*, 55 (3):147–150, 1948. → page 2

[5] M. A. Deakin. Yet more on urquhart's theorem. *The Australian Mathematical Society Gazette*, pages 3–3, 1997. → page 7

[6] D. Eppstein and D. Hart. Shortest paths in an arrangement with k line orientations. In *Symposium on Discrete Algorithms: Proceedings of the tenth annual ACM-SIAM symposium on Discrete algorithms*, volume 17, pages 310–316, 1999. → page 2

[7] A. Gajentaan and M. H. Overmars. On a class of $o(n^2)$ problems in computational geometry. *Computational geometry*, 5(3):165–185, 1995. → pages 6, 46

[8] M. Hajja. An elementary proof of the most "elementary" theorem of euclidean geometry. *J. Geometry Graphics*, 8:17–22, 2004. → pages 6, 7

[9] M. Hajja. A very short and simple proof of "the most elementary theorem" of euclidean geometry. In *Forum Geom*, volume 6, pages 167–169, 2006. → page 7

[10] D. Hart. Approximating the shortest path in line arrangements. In *CCCG*, pages 105–108. Citeseer, 2001. → pages 3, 24

[11] D. Hart. Shortest paths in two intersecting pencils of lines. In *CCCG*, pages 166–169, 2003. → pages 2, 7

[12] M. R. Henzinger, P. Klein, S. Rao, and S. Subramanian. Faster shortest-path algorithms for planar graphs. *Journal of Computer and System Sciences*, 55 (1):3–23, 1997. → page 2

[13] T. Kavitha and K. R. Varadarajan. On shortest paths in line arrangements. In *CCCG*, pages 170–173, 2003. → pages 2, 7

[14] J. King. A survey of 3SUM-hard problems. *Not published*, 2004. → pages 6, 46

[15] D. G. Kirkpatrick and R. Seidel. The ultimate planar convex hull algorithm? *SIAM journal on computing*, 15(1):287–299, 1986. → page 2

[16] A. Michael. Deakin. the provenance of urquhart's theorem. *Austral. Math. Soc. Gaz*, 8(1):26–28, 1981. → page 7

[17] D. Pedoe. The most "elementary" theorem of euclidean geometry. *Mathematics Magazine*, 49(1):40–42, 1976. → page 7