Page   **Discussion**   Read   Edit   View history

Search

Computational geometry

Main page
Index
Recent changes
Random page
Help

Tools

What links here
Related changes
Special pages
Printable version
Permanent link
Page information

Introduction and Scope [edit]

Definition [edit]

Computational geometry, as one can easily guess from the name, is the branch of computer science encompassing geometrical problems and how their solutions may be implemented (efficiently, one would hope) on a computer.

Scope   [[edit]()]

Essentially all of the computational geometry you will encounter in high-school level competitions, even competitions such as the IOI, is plane Euclidean geometry, the noble subject on which Euclid wrote his *Elements* and a favorite of mathematical competitions. You would be hard-pressed to find contests containing geometry problems in three dimensions or higher. You also do not need to worry about non-Euclidean geometries in which the angles of a triangle don't quite add to 180 degrees, and that sort of thing. In short, the type of geometry that shows up in computer science contests is the type of geometry to which you have been exposed, countless times, in mathematics class, perhaps without being told that other geometries exist. So in all that follows, the universe is two-dimensional, parallel lines never meet, the area of a circle with radius r is πr^2, and so on. This article discusses basic two-dimensional computational geometry; more advanced topics, such as computation of [convex hulls](), are discussed in separate articles.

Points   [[edit]()]

Introduction to points   [[edit]()]

Many would claim that the point is the fundamental unit of geometry. Lines, circles, and polygons are all merely (infinite) collections of points and in fact we will initially consider them as such in order to derive several important results. *A point is an exact location in space*. Operations on points are very easy to perform computationally, simply because points themselves are such simple objects.

Representation: Cartesian coordinates   [[edit]()]

In our case, "space" is actually a plane, two-dimensional. That means that we need two real

Are you a developer? Try out the [HTML to PDF API]()

numbers to describe any point in our space. Most of the time, we will be using the Cartesian (rectangular) coordinate system. In fact, when points are given in the input of a programming problem, they are almost always given in Cartesian coordinates. The Cartesian coordinate system is easy to understand and every pair of real numbers corresponds to exactly one point in the plane (and *vice versa*), making it an ideal choice for computational geometry.

Thus, a point will be represented in the computer's memory by *an ordered pair of real numbers*. Due to the nature of geometry, it is usually inappropriate to use integers, as points generally do not fit neatly into the integer lattice! Even if the input consists only of points with integer coordinates, calculations with these coordinates will often yield points with non-integral coordinates, which can often cause counter-intuitive behavior that will have you scratching your head! For example, in C++, when one integer is divided by another, the result is always truncated to fit into an integer, and this is usually not desirable.

Determining whether two points coincide [edit]

Using the useful property of the Cartesian coordinate system discussed above, we can determine whether or not two given points coincide. Denote the two points by P and Q, with coordinates (x_1, y_1) and (x_2, y_2), respectively, and then:

$$P = Q \longleftrightarrow x1 = x2 \land y1 = y2$$

Read: The statement that P and Q are the same point is equivalent to the statement that their corresponding coordinates are equal. That is, P and Q having both the same x-coordinates and also the same y-coordinates is both sufficient and necessary for the two to be the same point.

The distance between two points [edit]

To find the distance between two points, we use the Euclidean formula. Given two points

$P = (x_1, y_1)$ and $Q = (x_2, y_2)$, the distance between them is given by:

$$\text{dist}(P, Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

(Note that when we write $P = (x_1, y_1)$, we mean that P is a point with x-coordinate x_1 and y-coordinate y_1.)

The midpoint of the line segment joining two points   [edit]

Given two points, how may we find the midpoint of the line segment joining two points? (Intuitively, it is the point that is "right in the middle" of two given points. One might think that we require some knowledge about line segments in order to answer this, but it is for precisely the reason that, in a certain sense, no such knowledge is required to understand the answer, that this operation is found in the Points section. So, given two points $P = (x_1, y_1)$ and $Q = (x_2, y_2)$, this midpoint is given by

$$\text{midpoint}(P, Q) = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right)$$

Lines   [edit]

What is a line?   [edit]

In axiomatic geometry, some terms such as *point* and *line* are left undefined, because an infinite regression of definitions is clearly absurd. In the algebraic approach we are taking, the line is defined in terms of the points which lie on it; that will be discussed in the following section. It will just be pointed out here that the word *line* is being used in the modern mathematical sense. Lines are straight; the terms *line* and *straight line* shall have identical meaning. Lines extend indefinitely in

both directions, unlike *rays* and *line segments*.

The equation of a line [edit]

In computational geometry, we have to treat all aspects of geometry algebraically. Computers are excellent at dealing with numbers but have no mechanism for dealing with geometrical constructions; rather we must reduce them to algebra if we wish to accomplish anything.

In Ontario high schools, the *equation of a line* is taught in the ninth grade. For example, the line which passes through the points (0,1) and (1,0) has the equation $x + y = 1$. Precisely, this means that for a given point (x, y), the statement $x + y = 1$ is equivalent to, or sufficient and necessary for, the point to be on the line.

The form of the equation of the line which is first introduced is generally the $y = mx + b$, in which m is the slope of the line and b is the y-intercept. For example, the line discussed above has the equation $y = -x + 1$, that is, $m = -1$ and $b = 1$. By substituting different values for m and b, we can obtain various (different) lines. But there's a problem here: if your line is vertical, then it is not possible to choose values of m and b for the line. (Try it!) This is because the y-coordinate is no longer a function of the x-coordinate.

Thus, when dealing with lines computationally, it seems we would need to have a special case: check if the line is vertical; if so, then do something, otherwise do something else. This is a time-consuming and error-prone way of coding.

Standard Form of the equation of a line [edit]

Even though the slope-intercept form cannot describe a vertical line, there *is* an equation that describes a vertical line. For example, the line passing through (3,1) and (3,8) is $x = 3$. In fact,

almost any line can be described by an equation of the form $x = my + b$. (Try it if you don't believe me. I have merely switched around x and y from the slope-intercept form.) Except... *horizontal* lines. So we have two forms of the equation of a line: one which fails on vertical lines and one which fails on horizontal lines. Can we combine them to give an equation of the line which is valid for *any* line?

As it turns out, it is indeed possible.

That equation, the *standard form of the equation of the line* is:

$$Ax + By + C = 0$$

By substituting appropriate values of A, B, and C, one can describe any line with this equation. And by storing values of A, B, and C, one can represent a line in the computer's memory. Here are some pairs of points and possible equations for each:

$$(0, 1) \quad \text{and} \quad (1, 0): \quad x + y - 1 = 0$$
$$(3, 1) \quad \text{and} \quad (3, 8): \quad x - 3 = 0$$
$$(2, 5) \quad \text{and} \quad (4, 5): \quad y - 5 = 0$$
$$(6, 6) \quad \text{and} \quad (4, 9): \quad 3x + 2y - 30 = 0$$

As you can see, it handles vertical and horizontal lines properly, as well as lines which are neither.

Note that the standard form is not unique: for example, the equation of the first line could have just as well been $-x - y + 1 = 0$ or perhaps $5x + 5y - 5 = 0$. Any given line has infinitely many representations in the standard form. However, each standard form representation describes at most one line.

If A and B are both zero, the standard form describes no line at all.

Slope and intercepts of the line in standard form [edit]

By isolating y from the standard form, we obtain the slope and y-intercept form for line l ($Ax + By + C$) when $B \neq 0$ (that is, when the line is not vertical):

$$\text{slope}(l) = -\frac{A}{B}$$

The y-intercept is obtained by letting $x = 0$ and then:

$$\text{y-intercept}(l) = -\frac{C}{B}$$

Similarly, the x-intercept is given by:

$$\text{x-intercept}(l) = -\frac{C}{A}$$

In order for the x-intercept to exist, the line must not be horizontal, that is, $A \neq 0$.

Determining if a point is on a line [edit]

Using the definition of the equation of a line, it becomes evident that to determine whether or not a point lies on a line, we simply substitute its coordinates into the equation of the line, and check if the LHS is, indeed, equal to zero. This allows us to determine, for example, that $(12, -3)$ is on the last line given in the table above, whereas $(14, -7)$ is not on that line.

Construction of the line through two given points [edit]

A good question is: *how* do we determine that fourth equation above, the equation of the line through (6,6) and (4,9)? It's not immediately obvious from the two points given, whereas the other three are pretty easy.

For the slope-y-intercept form $y = mx + b$, you first determined the slope m, and then solved for b. A similar procedure can be used for standard form. We state here the following pseudocode for determining the coefficients A, B, C of the equation of the line through points (x_1, y_1) and (x_2, y_2) in standard form:

1. $A \leftarrow y_1 - y_2$
2. $B \leftarrow x_2 - x_1$
3. $C \leftarrow -Ax_1 - By_1$

(It is one thing to derive a formula or algorithm and quite another thing to prove it. The derivation of this formula is not shown, but proving it is as easy as substituting to determine that the line really does pass through the two given points.)

Parallel and coincident lines   [edit]

In the slope and y-intercept form, two lines are either parallel or coincident if they have the same slope. So given two lines $l_1 (A_1x + B_1y + C_1 = 0)$ and $l_2 (A_2x + B_2y + C_2 = 0)$, we obtain, for $B_1, B_2 \neq 0$:

$$-\frac{A_1}{B_1} = -\frac{A_2}{B_2}$$

Cross-multiplying gives a result that is valid even if either or both lines are vertical, that is, it is valid for any pair of lines in standard form:

$$\text{parallel\_or\_coincident}(l_1, l_2) \longleftrightarrow A_1 B_2 = A_2 B_1$$

Now, in the slope-y-intercept form, two lines coincide if their slopes and y-intercepts both coincide. In a manner similar to that of the previous section, we obtain:

$$B_1 C_2 = B_2 C_1$$

or, if x-intercepts are used instead:

$$A_1 C_2 = A_2 C_1$$

Two lines coincide if $A_1 B_2 = A_2 B_1$ *and* either of the two equations above holds. (As a matter of fact, if the two lines are coincident, they will both hold, but only one of them needs to be checked.) If $A_1 B_2 = A_2 B_1$ but the lines are not coincident, they are parallel.

Finding the point of intersection of two lines

If two lines are coincident, every point on either line is an intersection point. If they are parallel, then no intersection points exist. We consider the general case in which neither is true.

In general, two lines intersect at a single point. That is, the intersection point is the single point that lies on both lines. Since it lies on both lines, it must satisfy the equations of both lines simultaneously. So to find the intersection point of $l_1 \left(A_1 x + B_1 y + C_1 = 0 \right)$ and l_2 $\left(A_2 x + B_2 y + C_2 = 0 \right)$, we seek the ordered pair $\left(x, y \right)$ which satisfies:

$$\begin{aligned} A_1 x + B_1 y + C_1 &= 0 \\ A_2 x + B_2 y + C_2 &= 0 \end{aligned}$$

This is a system of two linear equations in two unknowns and is solved by Gaussian elimination. Multiply the first by A_2 and the second by A_1, giving:

$$\begin{aligned} A_1 A_2 x + A_2 B_1 y + A_2 C_1 &= 0 \\ A_1 A_2 x + A_1 B_2 y + A_1 C_2 &= 0 \end{aligned}$$

Subtracting the former from the latter gives, with cancellation of the x term:

$$\begin{aligned}
(A_1B_2 - A_2B_1)y + (A_1C_2 - A_2C_1) &= 0 \\
(A_1B_2 - A_2B_1)y &= A_2C_1 - A_1C_2 \\
y &= \frac{A_2C_1 - A_1C_2}{A_1B_2 - A_2B_1}
\end{aligned}$$

Instead of redoing this to obtain the value of x, we take advantage of symmetry and simply swap all the A's with the B's. (If you don't believe that this works, do the derivation the long way.)

$$x = \frac{B_2C_1 - B_1C_2}{B_1A_2 - B_2A_1}$$

or:

$$x = \frac{B_1C_2 - B_2C_1}{A_1B_2 - A_2B_1}$$

Notice the quantity $A_1B_2 - A_2B_1$, and how it forms the denominator of the expressions for both x and y. When solving for the intersection point on the computer, you only need to calculate this quantity once. (This quantity, called a *determinant*, will resurface later on.) Here is pseudocode for finding the intersection point:

1. $det \leftarrow A_1B_2 - A_2B_1$
2. if $det = 0$
3.   fail
4. else
5. $x \leftarrow (B_1C_2 - B_2C_1)/det$
6. $y \leftarrow (A_2C_1 - A_1C_2)/det$

When $det = 0$, the lines are either parallel or coincident. We now see algebraically that the division by zero prevents us from finding a unique intersection point for such pairs of lines.

Direction numbers for a line   [edit]

This in itself is not very useful, but it will become important in the following sections as a simplifying concept.

Lines are straight; effectively they always point in the same direction. One way to express that direction has been slope, which unfortunately is undefined for vertical lines. *The slope m for a line told us that you could* start at any point on the line, move Δx units to the right, then move $m \Delta x$ units up, and you would again be located on the line. Thus we can say that $(1, m)$ is a pair of direction numbers for that line. This means that if (x_0, y_0) is on a line, and Δx and Δy are in the ratio $1 : m$, for that line, then $(x_0 + \Delta x, y_0 + \Delta y)$ is on the same line. This means that $(2, 2m)$ is also a set of direction numbers for that line, or, indeed, any multiple of $(1, m)$ other than $(0, 0)$. ($(0, 0)$ clearly tells you nothing about the line.)

We can define something similar for the line in standard form. Choose some starting point (x_0, y_0) on line l. Now, move to a new point $(x_0 + \Delta x, y_0 + \Delta y)$. In order for this point to be on the line l, we must have

$$A(x_0 + \Delta x) + B(y_0 + \Delta y) + C = 0$$

Expanding and rearranging gives

$$A x_0 + B y_0 + C + A \Delta x + B \Delta y = 0$$

We know that $A x_0 + B y_0 + C = 0$ since (x, y) is on line l. Therefore,

$$\begin{aligned} A \Delta x + B \Delta y &= 0 \\ A \Delta x &= -B \Delta y \end{aligned}$$

Convince yourself, by examining the equation above, that $(-B, A)$ is a set of direction numbers

for line l. Similarly, if we have a pair of direction numbers $(\Delta x, \Delta y)$, although this does not define a unique line, we can obtain possible values of A and B as $-\Delta y$ and Δx, respectively. The relationship between direction numbers and points on the corresponding line is an "if-and-only-if" relationship. If Δx and Δy are in the ratio $-B : A$, then we can "shift" by $(\Delta x, \Delta y)$, and *vice versa*.

Any line perpendicular to l will have the direction numbers (A, B), and thus a possible equation starts $-Bx + Ay + \ldots = 0$. (This is the same as saying, for non-vertical lines, that the product of slopes of perpendicular lines is -1. Examine the equation for the slope of a line given in standard form and you'll see why.) In fact, in an algebraic treatment of geometry such as this, we do not prove this claim, but instead proclaim it the definition of perpendicularity: given two lines with direction numbers (A, B) and (C, D), they are perpendicular if and only if $\mathbb{E}$ $AC + BD = 0$.

Given some line, all lines parallel to that one have the same direction numbers. That is, the direction numbers, while providing information about a line's direction, provide no information about its position. However, sometimes all that is needed is the direction, and here the direction numbers are very useful.

Parametric equation of a line   [edit]

Given some line $Ax + By + C = 0$ and a point (x_0, y_0) known to be on the line, we can use the direction numbers to write the parametric equation of the line. Since $(-B, A)$ is a set of direction numbers for the line, we see that $(x_0 - Bt, y_0 + At)$ will be on the line for any t, and furthermore each point on the line corresponds to exactly one value of t in this way.

Translation along a line [edit]

Given some line $Ax + By + C = 0$ and a point (x_0, y_0) known to be on the line, we wish to locate another point on the line that is distance h away from (x_0, y_0). We can do this by observing that the general point is given by $(x_0 - Bt, y_0 + At)$, which is distance

$$\sqrt{(Bt)^2 + (At)^2} = t\sqrt{A^2 + B^2} = h \ \text{away from} \ (x_0, y_0). \text{Therefore, we want}$$

$t = \dfrac{h}{\sqrt{A^2 + B^2}}$, and finally obtain $\left(x_0 - \dfrac{Bh}{\sqrt{A^2 + B^2}}, y_0 + \dfrac{Ah}{\sqrt{A^2 + B^2}}\right)$. Note that

unless $h = 0$, there is another possible answer, which is simply "on the other side":

$$\left(x_0 + \dfrac{Bh}{\sqrt{A^2 + B^2}}, y_0 - \dfrac{Ah}{\sqrt{A^2 + B^2}}\right).$$

Dropping a perpendicular [edit]

Given a line $l\,(Ax + By + C = 0)$ and a point $P\,(x_0, y_0)$ which may or may not be on l, can we find the line perpendicular to l passing through P? By Euclid's Fifth Postulate, there exists exactly one such line. The algorithm to find it is given below:

1. $A' \leftarrow -B$
2. $B' \leftarrow A$
3. $C' \leftarrow Bx_0 - Ay_0$

where $A'x + B'y + C' = 0$ is the perpendicular line desired.

There is nothing difficult to memorize here: we already noted in the previous section how to find the values of A' and B', and finding the value of C' is merely setting $A'x_0 + B'y_0 + C'$

equal to zero (so that the point (x_0, y_0) will be on the resulting line).

The *foot* of the perpendicular is the point at which it intersects the line l. It is guaranteed to exist since two lines cannot, of course, be both perpendicular and parallel. Combining the above algorithm with the line intersection algorithm explained earlier gives a solution for the location of the point. A bit of algebra gives this optimized algorithm:

1. $z \leftarrow \frac{Ax_0 + By_0 + C}{A^2 + B^2}$
2. $x \leftarrow x_0 - Az$
3. $y \leftarrow y_0 - Bz$

where (x, y) are the coordinates of the foot of the perpendicular from P to l.

The distance from a point to a line   [edit]

By the distance from a point $P\,(x_0, y_0)$ to a line $l\,(Ax + By + C = 0)$ what is meant is the closest possible distance from P to any point on l. What point on l is closest to P? It is intuitive perhaps that it is obtained by dropping a perpendicular from P to l. That is, we choose a point Q such that $PQ \perp l$, and the distance from P to l is then the length of line segment $\overline{PQ}$, denoted $|PQ|$.

The reason why this is the shortest distance possible is this: Choose any other point R on l. Now, $\triangle PQR$ is right-angled at Q. The longest side of a right triangle is the hypotenuse, so that $|PR| > |PQ|$. Thus $|PQ|$ is truly the shortest possible distance.

Now, as noted earlier, the line PQ, being perpendicular to l, has the direction numbers (A, B). Thus, for any t, the point $(x_0 + At, y_0 + Bt)$ is on PQ. For some choice of t, this point must coincide with Q. Since that point lies on l, we have

$$\begin{aligned}
A(x_0 + At) + B(y_0 + Bt) + C &= 0 \\
Ax_0 + By_0 + C + A^2t + B^2t &= 0 \\
(A^2 + B^2)t &= -(Ax_0 + By_0 + C) \\
t &= -\frac{Ax_0 + By_0 + C}{A^2 + B^2}
\end{aligned}$$

This instantly gives the formula for the foot of the perpendicular given in the previous section.

Now, the distance $|PQ|$ is found with the Euclidean formula:

$$\begin{aligned}
|PQ| &= \sqrt{(At)^2 + (Bt)^2} \\
&= \sqrt{A^2t^2 + B^2t^2} \\
&= \sqrt{(A^2 + B^2)t^2} \\
&= |t|\sqrt{A^2 + B^2} \\
&= \frac{|Ax_0 + By_0 + C|}{A^2 + B^2}\sqrt{A^2 + B^2} \\
&= \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}
\end{aligned}$$

The last line is the formula to remember. To restate,

$$\text{dist}(P, l) = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}$$

(It was noted earlier that if A and B are both zero, then we don't actually have a line. Therefore, the denominator above can never be zero, which is a good thing.)

On which side of a line does a point lie? [edit]

A line partitions the plane into two regions. For example, a vertical line divides the plane into a region on the left and a region on the right. Now, when a point $P(x_0, y_0)$ does not satisfy the

equation of a line l $(Ax + By + C = 0)$, can we determine on which side of the line it lies?

Yes we can, with a certain restriction. If $Ax_0 + By_0 + C > 0$, then the point lies on one side of the line; if $Ax_0 + By_0 + C < 0$, then it lies on the other side. However, it's a bit pointless to say *which* side it lies on: does it lie on the left or the right? If the line is horizontal, then this question becomes meaningless. Also, notice that if we flip the signs of A, B, and C, then the value of $Ax_0 + By_0 + C$ is negated also, but that changes neither the point or the line. It is enough, however, to tell if two points are on the same side of the line or on opposite sides; simply determine whether $Ax + By + C$ has the same sign for both, or different signs.

The distance between two lines   [edit]

If two lines intersect, the closest distance between them is zero, namely at their intersection point. If they are coincident, then the distance is similarly zero. If two lines are parallel, however, there is a nonzero distance between them, and it is defined similarly to the distance between a point and a line. To find this distance, we notice that the lines OP and OQ are coincident, where O is the origin, P is the foot of the perpendicular from O to l_1, and Q is the foot of the perpendicular from O to l_2. We know, by substituting $(x_0, y_0) = (0, 0)$ for both l_1 ($A_1 x + B_1 y + C_1 = 0$) and l_2 ($A_2 x + B_2 y + C_2 = 0$) that:

$$|OP| = \frac{C_1}{\sqrt{A_1^2 + B_1^2}}$$
$$|OQ| = \frac{C_2}{\sqrt{A_2^2 + B_2^2}}$$

Now here we come up against a complication. If O is on the line *segment* $\overline{PQ}$, then we have to *add* $|OP|$ and $|OQ|$ to get the desired $|PQ|$. That is, if O is on the same side of both lines. Otherwise, we have to take the difference. Here's some code that takes care of these details:

1. $d_1 \leftarrow C_1/\sqrt{A_1^2 + B_1^2}$
2. $d_2 \leftarrow C_2/\sqrt{A_2^2 + B_2^2}$
3. if $A_1 \neq 0$
4.   if A_1 and A_2 have the same sign
5. $dist \leftarrow |d_1 - d_2|$
6.   else
7. $dist \leftarrow |d_1 + d_2|$
8. else :
9.   if B_1 and B_2 have the same sign
10. $dist \leftarrow |d_1 - d_2|$
11.   else
12. $dist \leftarrow |d_1 + d_2|$

Translating a line [edit]

Suppose we are given a line $Ax + By + C = 0$ and we wish to find a line parallel to it such that the distance between these two lines is h. To do this, we suppose that the new line has equation $Ax + By + D = 0$. Since the two lines have the same A and B value, they must be parallel (or coincident of $C = D$ also). Then, the distance between them will be given by $\dfrac{|C - D|}{\sqrt{A^2 + B^2}}$, as the previous section suggests. To obtain a value of h for this expression, it is not hard to see that we should choose $D = C \pm h\sqrt{A^2 + B^2}$.

Line segments [edit]

Are you a developer? Try out the HTML to PDF API

Introduction to line segments   [edit]

A line segment is the part of a line located "between" two points on that line, called *endpoints*. Any pair of points defines a unique line segment. Most definitions of "line segment" allow the endpoints to coincide, giving a single point, but this case will often not arise in programming problems and it is trivial to handle when it does arise, so we will not discuss it here; we assume the endpoints must be distinct. Thus, every line segment defines exactly one line.

We may represent a line segment in memory as a pair of points: that is, four numbers in total.

Coincidence (equivalence) of line segments   [edit]

Two line segments coincide if they have the same endpoints. However, they may have them in any order, hence we have:

$$\overline{PQ} = \overline{RS} \longleftrightarrow (P = R \vee Q = S) \wedge (P = S \vee Q = R)$$

(The parentheses are unnecessary and are added only for the sake of clarity.)

Length of a line segment   [edit]

The length of a line segment is nothing more than the distance between its endpoints.

Partitioning by length   [edit]

Suppose we wish to partition a line segment $\overline{PQ}$ by introducing a point R on $\overline{PQ}$ such that $|PR|/|RQ| = r : s$. That is, we wish to partition it into two line segments with their lengths in the ratio r/s. We may do so as follows:

$$\text{partition}(\overline{PQ}, r, s) = \left(\frac{sP_x + rQ_x}{s + r}, \frac{sP_y + rQ_y}{s + r} \right)$$

In the special case that $r = s$, we have the midpoint, as discussed in the Points section.

Containing line [edit]

All we have to do is find the line passing through both endpoints; the algorithm to do this is discussed in the section "Construction of the line through two given points".

Determining if a point lies on a line segment [edit]

Here is one interesting idea: if a point R lies on segment $\overline{PQ}$, then the relation $|PR| + |RQ| = |PQ|$ will hold. If it is on the *line* PQ but not on the *segment* $\overline{PQ}$, then it is the *difference* between $|PR|$ and $|RQ|$ that will equal $|PQ|$, not the sum. If it is not on this line, then the points P, Q, and R form a triangle, and by the Triangle Inequality, $|PR| + |RQ| > |PQ|$. Thus:

$$\text{on\_segment}(R, \overline{PQ}) \longleftrightarrow \text{dist}(P, R) + \text{dist}(R, Q) = \text{dist}(P, Q)$$

Although this test is mathematically ingenious, it should not be used in practice, since the extraction of a square root is a slow operation. (Think about how much work it takes by hand, for example, to compute a square root, relative to carrying out multiplication or division by hand.) A faster method is to obtain the line containing the line segment (see previous section); if multiple queries are to be made on the same line segment then it is advisable to store the values of A, B, and C rather than computing them over and over again; and we first check if the point to test is on the *line*; if it is, then we must check if it is on the *segment* by checking if each coordinate of the point is between the corresponding coordinates of the endpoints of the segment.

For a one-time query (when we do not expect to see the line again), the use of the properties of similar triangles yields the following test: the point is on the line segment if and only if $(R_x - P_x)(Q_y - R_y) = (Q_x - R_x)(R_y - P_y)$ and R is between P and Q. When this test is used several times with different segments, the number of multiplications required is only half of the number required for the test *via* the containing line, but if the line is reused, then the test *via* the containing line ends up using fewer additions/subtractions in the long run.

Intersection of line segments   [edit]

Given two line segments $\overline{PQ}$ and $\overline{RS}$, how do we determine whether they intersect?

First, if the containing lines are coincident, then the line segments intersect if and only if at least one of the endpoints of one of the segments is on the other segment. In general, when the containing lines do not coincide, the segments intersect if and only if P and Q are not on the same side of RS *and* R and S are not on the same side of PQ. That is, extend each segment to a line and then determine on which sides of the line lie the endpoints of the other segment. (If one point is on the line and the other is not, then they are not considered to be on the same side, since two line segments can intersect even if the endpoint of one lies on the other.)

If the segments intersect, their intersection point can be determined by finding the intersection point of the containing lines.

Another method for determining whether two line segments intersect is finding the intersection point (if it exists) of the containing lines and checking if it lies on both line segments (as in the end of the previous section). After finding the containing lines, this method requires six multiplications and two divisions, whereas the one above requires eight multiplications. Since multiplications are generally faster, we prefer the method above to this one.

Do two line segments cross?   [edit]

The word *cross* is used here in a stronger sense than *intersect*. Two line segments cross if they intersect and no point of intersection is an endpoint of either line segment. Intuitively, the two line segments form a (possibly distorted) X shape. Here, we can ignore the degenerate cases for line segment *intersection*: we simply test that P and Q are on different sides of RS (this time the test *fails* if either of them is actually *on* it), and that R and S are on different sides of PQ. The second method described in the previous section can again be applied, although again it is expected to be slower.

Direction numbers for the containing line [edit]

By the definition of the direction numbers, a set of direction numbers for the line segment $\overline{PQ}$ is $\left(Q_x - P_x, Q_y - P_y\right)$. This gives an instant proof for the "magic formula" for the line through two given points: we convert these direction numbers to values for A and B and then solve for C using one of the points.

Perpendicular bisector of a line segment [edit]

The perpendicular bisector of a line segment is the line perpendicular to the line segment which also passes through the line segment's midpoint. Notice that the direction numbers obtained in the previous section can be used to obtain the direction numbers for a perpendicular line, and that these can in turn be used to reconstruct the values of A and B for that line. Given that the line must also pass through the midpoint:

1. $A \leftarrow Q_x - P_x$
2. $B \leftarrow Q_y - P_y$
3. $x \leftarrow (Q_x + P_x)/2$
4. $y \leftarrow (Q_y + P_y)/2$
5. $C \leftarrow -Ax - By$

This technique requires a total of two multiplications and two divisions. If we substitute the values of x and y into the last line and expand, we can change this to four multiplications and one division, which is almost certainly slower as division by two is a very fast operation.

Conclusion - lines and line segments [edit]

The techniques of the two preceding chapters should provide inspiration on how to achieve tasks that are "somewhere in-between". For example, we have not discussed the intersection of a line and a line segment. However, it is fairly clear that all that is required is to test on which sides of the line lie the endpoints of the segment: half of the test for two line segments. We also have not discussed the distance from a point to a line segment. We have omitted any discussion of rays altogether. If you thoroughly understand how these techniques work, though, extending them to problems not explicitly mentioned should not be difficult. Feel free to add these sections to this article; the exclusion of any material from the current draft is not an indication that such material does not belong in this article.

Angles [edit]

Introduction to angles [edit]

In trigonometry, one proves the Law of Cosines. In a purely algebraic approach to geometry,

however the concept of angle is *defined* using the Law of Cosines, and the Law itself requires no proof. Still, we will not use that definition directly, because deriving everything from it would be unnecessarily complicated. Instead, we will assume that we already know some properties of angles. Storing angles in memory is very easy: just store the angle's radian measure. Why not degree measure? Degree measure is convenient for mental calculation, but radian measure is more mathematically convenient and, as such, trigonometric functions of most standard language libraries, such as those of Pascal and C/C++, expect their arguments in radians. (Inverse trigonometric functions return results in radians.) We will use radian measure throughout this chapter, without stating "radians", because radian measure is assumed when no units are given.

Straightforward applications of basic trigonometry, such as finding the angles in a triangle whose vertices are known (Law of Cosines), are not discussed here.

Directed angle and the atan2 function   [edit]

Suppose a ray with its endpoint at the origin initially points along the positive x-axis and is rotated counterclockwise around the origin by an angle of θ. From elementary trigonometry, the ray now consists of points $(k\cos\theta, k\sin\theta)$, where $k > 0$. This angle θ is a *directed* angle. Notice there is another ray with its endpoint at the origin that makes an angle of θ with the positive x-axis: obtained by rotating clockwise rather than counterclockwise. But the directed angle in this case would be $-\theta$. Thus, by specifying a directed angle from the positive x-axis we can uniquely specify one particular ray.

Can we reverse this process? Can we find the directed angle from the positive x-axis to the ray $\overrightarrow{OP}$, where $P = (x, y)$? Notice that when $x \neq 0$, $y/x = \tan\theta$, so taking the inverse tangent should give back θ. There are just two problems with this: one is that x might be zero (but the angle will still be defined, either $\pi/2$ or $3\pi/2$), the other is that the point $(-x, -y)$ will give

the same tangent even though it lies on the other side (and hence its directed angle should differ from that of $\overrightarrow{OP}$ by π. However, because this is such a useful application, the Intel FPU has a built-in instruction to compute the directed angle from the positive x-axis to ray $\overrightarrow{OP}$, and the libraries of both C and Free Pascal contain functions for this purpose. C's is called `atan2`, and it takes two arguments, y and x, in that order, returning an angle in radians, the desired directed angle, a real number θ satisfying $-\pi < \theta \leq \pi$. (Notice that as with undirected angles, adding 2π to a directed angle leaves it unchanged). Remember that y comes first and not x; the reason for this has to do with the design of the Intel FPU and the calling convention of C. Free Pascal's math library aims to largely emulate that of C, so it provides the `arctan2` function which takes the same arguments and produces the same return value.

The angle between a line and the x-axis   [edit]

When two lines intersect, two pairs of angles are formed (the two angles in each pair are equal). They are supplementary. To find one of these angles, let us shift the line until it passes through the x-axis. Then, adding the direction numbers $(-B, A)$ to the origin gives another point. We now apply the `atan2` function: `atan2`$(A, -B)$. (Notice that we have reversed the order of x and y, as required.) The result may be negative; we can add π to it to make it non-negative.

The angle between two lines   [edit]

To find one of the two angles between two lines, we find the angle between each line and the x-axis, then subtract. (Draw a diagram to convince yourself that this works.) If the result is negative, add π degrees, once or twice as necessary. The other angle is obtained by subtracting from π.

The angle bisector of a pair of intersecting lines   [edit]

Using the result of the previous section and a great deal of algebra and trigonometry, together with the line intersection algorithm, gives the following algorithm for finding one of the two angle bisectors of a pair of (intersecting) lines $l_1 \, (A_1 x + B_1 y + C_1 = 0)$ and l_2 $(A_2 x + B_2 y + C_2 = 0)$ (the bisector is represented as $A'x + B'y + C' = 0$):

1. $h_1 \leftarrow \sqrt{A_1^2 + B_1^2}$
2. $h_2 \leftarrow \sqrt{A_2^2 + B_2^2}$
3. $A' \leftarrow A_1 h_2 + A_2 h_1$
4. $B' \leftarrow B_1 h_2 + B_2 h_1$
5. $(x, y) \leftarrow$ intersection of l_1 and l_2
6. $C' \leftarrow -A'x - B'y$

The other angle bisector, of course, is perpendicular to this line and also passes through that intersection point.

Vectors   [edit]

Introduction to vectors   [edit]

Although the word *vector* has a formal definition, we will not find it useful. It is better, in the context of computational geometry, to think of a vector as an idealized object which represents a given translation. Visually, a vector may be represented as an arrow with a fixed length pointing in a fixed direction, but without a fixed location. For example, on the Cartesian plane, consider the translation "3 units down and 4 units right". If you start at the point (8,10) and apply this translation, (12,7) is obtained. This can also be represented as an arrow with length 5 units and direction approximately 37 degrees south of east. Placing the tail of this arrow at (8,10) results in the head resting upon the point (12,7). The vector from (0,0) to (4,-3) is the same vector, as it also

Are you a developer? Try out the HTML to PDF API

represents a translation 3 units down and 4 units right, and has the same length and direction. However, the vector from (12,7) to (8,10) and the vector from (0,0) to (-3,4) are different from the one from (8,10) to (12,7); they have the same lengths, but their directions are different, and so they are different vectors. The vector from (0,0) to (8,-6) is also different; it has the same direction, but a different length. In both cases, these are simply different translations, and hence different vectors. The word *vector* is from the Latin, meaning *carrier*, appropriate as it *carries* from one location (a point) to another.

Representation   [edit]

We could represent a vector by its magnitude and direction, but it is usually not convenient to do so. Even still, it might be useful occasionally to find the magnitude and direction of a vector; refer to the sections on [length of a line segment](#) and [directed angle and the atan2 function](#) for the necessary mathematics. Instead, we will use the Cartesian representation. If we place the tail of a vector on the origin, the head rests upon a certain point; the Cartesian *coordinates* of this point are the Cartesian *components* of the vector. We thus represent a vector as we do a point, as an ordered pair of real numbers, but we shall enclose these in brackets instead of parentheses. The vector discussed above is then [4,-3], as when the tail is placed on (0,0), the head rests upon (4,-3). Two vectors are equal when both of their corresponding components are equal.

Standard vector notation   [edit]

In writing, vectors are represented as letters with small rightward-pointing arrows over them, *e.g.*, $\vec{v}$. Vectors are usually *typeset* in bold, *e.g.*, $\mathbf{v}$. We shall denote the components of $\mathbf{v}$ as v_x and v_y. The magnitude of $\mathbf{v}$ may be denoted simply v, or, to avoid confusion with an unrelated scalar variable, $\|\mathbf{v}\|$. There is a special vector known as the *zero vector*, the identity translation; it leaves

one's position unchanged and has the Cartesian representation [0,0]. We shall represent it as $\mathbf{0}$.

Basic operations [edit]

The order of operations is the same with vectors as with real numbers.

Addition [edit]

Addition of vectors shares many properties with addition of real numbers. It is represented by the plus sign, and is associative and commutative.

Of a vector to a point [edit]

Adding a vector to a point is another name for the operation in which a point is simply translated by a given vector. It is not hard to see that $(x, y) + [u, v] = (x + u, y + v)$. We could also write the same sum as $[u, v] + (x, y)$, since addition is supposed to be commutative. As one can easily see, adding a point and a vector results in another point. The zero vector is the additive identity; adding it to any point leaves the point unchanged. Geometrically, adding a point and a vector entails placing the vector's tail at the point; the vector's head then rests upon the sum.

Of two vectors [edit]

If we want addition to be associative too, then the sum $(P + \mathbf{u}) + \mathbf{v}$ must be the same point as $P + (\mathbf{u} + \mathbf{v})$. The first sum represents the point arrived at when the translations representing $\mathbf{u}$ and $\mathbf{v}$ are taken in succession. Therefore, in the second sum, we should define $\mathbf{u} + \mathbf{v}$ in such a way so that it gives a vector representing the translation obtained by taking those of $\mathbf{u}$ and $\mathbf{v}$ in succession. It is not too hard to see that $[u_x, u_y] + [v_x, v_y] = [u_x + v_x, u_y + v_y]$. Geometrically, if the tail of $\mathbf{v}$ is placed at the head of $\mathbf{u}$, the arrow drawn from the tail of $\mathbf{u}$ to the head of $\mathbf{v}$ is the sum. Again, the zero vector is the additive identity.

Negation   [edit]

For every vector $\mathbf{v}$ we can identify a corresponding vector, denoted $-\mathbf{v}$, such that the translations represented by $\mathbf{v}$ and $-\mathbf{v}$ are inverse transformations. Another way of stating this is that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$, or that $\mathbf{v}$ and $-\mathbf{v}$ have the same length but opposite directions. If $\mathbf{v} = [v_x, v_y]$, then $-\mathbf{v} = [-v_x, -v_y]$.

Subtraction   [edit]

Of a vector from a point   [edit]

We define subtraction to be the inverse operation of addition. That is, for any point P and vector $\mathbf{v}$, we should have that $P + \mathbf{v} - \mathbf{v} = P - \mathbf{v} + \mathbf{v} = P$. Since the vector $-\mathbf{v}$ represents the inverse translation to $\mathbf{v}$, we can simply **add** $-\mathbf{v}$ to P to obtain $P - \mathbf{v}$. (Note that the expression $\mathbf{v} - P$ is not meaningful.) Geometrically, this is equivalent to placing the head of the arrow at P and then following the arrow backward to the tail. Subtracting the zero vector leaves a point unchanged.

Of a vector from a vector   [edit]

To subtract one vector from another, we add its negative. Again, we find that this definition leads to subtraction being the inverse operation of addition. It is not too hard to see that $[u_x, u_y] - [v_x, v_y] = [u_x - v_x, u_y - v_y]$. Geometrically, if the vectors $\mathbf{u}$ and $\mathbf{v}$ are placed tail-to-tail, then the vector from the head of $\mathbf{u}$ to the head of $\mathbf{v}$ is $\mathbf{v} - \mathbf{u}$, and *vice versa*. Note that $\mathbf{u} - \mathbf{v} = -(\mathbf{v} - \mathbf{u})$.

Of a point from a point   [edit]

A vector can be considered the difference between two points, or the translation required to take one point onto the other. Given the two points $P\ (P_x, P_y)$ and $Q\ (Q_x, Q_y)$, the vector from P to Q is $[Q_x - P_x, Q_y - P_y]$, and *vice versa*. Note that $P - Q = -(Q - P)$.

Multiplication   [edit]

A vector can be *scaled* by a scalar (real number). This operation is known as *scalar multiplication*. The product $\alpha[v_x, v_y]$ denotes the vector $[\alpha v_x, \alpha v_y]$. This could also be notated $[v_x, v_y]\alpha$, though placing the scalar after the vector is more rare. Scalar multiplication takes precedence over addition and subtraction. The geometric interpretation is a bit tricky. If $\alpha > 0$, then the direction of the vector is left unchanged and the length is scaled by the factor α. If $\alpha = 0$, then the result is the zero vector; and if $\alpha < 0$, then the vector is scaled by the factor $|\alpha|$ and its direction is reversed. The scalar 1 is the multiplicative identity. Multiplying by the scalar -1 yields the negative of the original vector. Scalar multiplication is distributive.

The zero vector is a scalar multiple of any other vector. However, if two non-zero vectors are multiples of each other, then either they are *parallel*, that is, in the same direction ($\alpha > 0$), or they are *antiparallel*, that is, in opposite directions ($\alpha < 0$).

Division   [edit]

A vector can be divided by a scalar α by multiplying it by α^{-1}. This is notated in the same way as division with real numbers. Hence, $\dfrac{[x, y]}{\alpha} = \left[\dfrac{x}{\alpha}, \dfrac{y}{\alpha}\right]$. Division by 0 is illegal.

The unit vector   [edit]

To every vector except $\mathbf{v}$ (except $\mathbf{v} = \mathbf{0}$), we can assign a vector $\hat{\mathbf{v}}$ that has the same direction

but a length of exactly 1. This is called a *unit vector*. The unit vector can be calculated as follows:

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

There are two *elementary unit vectors*: [1,0], denoted $\hat{\boldsymbol{\imath}}$, and [0,1], denoted $\hat{\boldsymbol{\jmath}}$. That is, the unit vectors pointing along the positive x- and y-axes. Note that we can write a vector $\mathbf{v}$ as $v_x\hat{\boldsymbol{\imath}} + v_y\hat{\boldsymbol{\jmath}}$.

Obtaining a vector of a given length in the same direction as a given vector   [edit]

Suppose we want a vector of length l pointing in the same direction as $\mathbf{v}$. Then, all we need to do is scale $\mathbf{v}$ by the factor $\dfrac{l}{\|\mathbf{v}\|}$. Thus our new vector is $\dfrac{l}{\|\mathbf{v}\|}\mathbf{v}$, which can also be written $l\,\dfrac{\mathbf{v}}{\|\mathbf{v}\|} = l\hat{\mathbf{v}}$. Hence the unit vector is a "prototype" for vectors of a given direction.

Rotation   [edit]

Given a vector $\mathbf{v}$ $[x, y]$ and an angle θ, we can rotate $\mathbf{v}$ counterclockwise through the angle θ to obtain a new vector $[x', y']$ using the following formulae:

$$x' = x\cos\theta - y\sin\theta$$
$$y' = x\sin\theta + y\cos\theta$$

The same formula can be used to rotate points, when they are considered as the endpoints of vectors with their tails at the origin.

Dot product   [edit]

The *dot product* or *scalar product* is notated and defined as follows:

$$[a_x, a_y] \cdot [b_x, b_y] = a_x b_x + a_y b_y$$

It is commutative, but not associative (since expressions like $\mathbf{a} \cdot \mathbf{b} \cdot \mathbf{c}$ are meaningless). However, the dot product does associate with scalar multiplication; that is, $\alpha(\mathbf{a} \cdot \mathbf{b}) = (\alpha\mathbf{a}) \cdot \mathbf{b}$. It is distributive over addition, so that $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$. (This is actually left-distributivity, but right-distributivity follows from the commutative property.) These properties can easily be proven using the Cartesian components. The dot product has two useful properties.

First, the dot product satisfies the relation

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos\theta$$

where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. This allows us to find **the angle between two vectors** as follows:

$$\theta = \cos^{-1} \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$$

This formula breaks down if one of the vectors is the zero vector (in which case no meaningful angle can be defined anyway). The dot product gives a quick test for perpendicularity: two nonzero vectors are perpendicular if and only if their dot product is zero. The dot product is maximal when the two vectors are parallel, and minimal (maximally negative) when the two vectors are antiparallel.

Second, the dot product can be used to compute *projections*.

The vector projection   [edit]

The vector projection of $\mathbf{a}$ onto $\mathbf{b}$ is, intuitively, the "shadow" cast by $\mathbf{a}$ onto $\mathbf{b}$ by a light source delivering rays perpendicular to $\mathbf{b}$. We imagine $\mathbf{b}$ to be a screen and $\mathbf{a}$ to be an arrow; we are *projecting* the arrow onto the screen. (It is okay for parts of the projection to lie outside the "screen".) Geometrically, the tail of the projection of $\mathbf{a}$ onto $\mathbf{b}$ is the foot of the perpendicular from the tail of $\mathbf{a}$ to $\mathbf{b}$, and the head of the projection is likewise the foot of the perpendicular from the head of $\mathbf{a}$. The vector projection is a vector pointing in the same direction as $\mathbf{b}$. There is no standard notation for vector projection; one notation is $\mathrm{proj}_{\mathbf{b}}\, \mathbf{a}$ for the projection of $\mathbf{a}$ onto $\mathbf{b}$, and *vice versa*.

The scalar projection   [edit]

The scalar projection is the directed length of the vector projection. That is, if the angle between $\mathbf{a}$ and $\mathbf{b}$ is acute, then $\mathrm{proj}_{\mathbf{b}}\, \mathbf{a}$ points in the same direction as $\mathbf{b}$, and it has a positive directed length; here the scalar projection simply equals the length of the vector projection. If that angle is obtuse, on the other hand, then $\mathrm{proj}_{\mathbf{b}}\, \mathbf{a}$ and $\mathbf{b}$ point in opposite directions, and the scalar projection is the negative of the length of the vector projection. (If the two vectors are perpendicular, the scalar projection is zero.) There is also no standard notation for the scalar projection; one possibility is $\left| \mathrm{proj}_{\mathbf{b}}\, \mathbf{a} \right|$. (Note the single vertical bars, instead of the double vertical bars that denote length.) You have already encountered scalar projections: in the unit circle, the sine of an angle is the scalar projection of a ray making that directed angle with the positive x-axis onto the y-axis, and likewise the cosine is a scalar projection onto the x-axis.

Computing projections   [edit]

Why have we left discussion of the computation of the scalar and vector projections out of their respective sections? The answer is that the scalar projection is easier to compute, but harder to explain. Place the vectors $\mathbf{a}$ and $\mathbf{b}$ tail-to-tail at the origin. Now rotate both vectors so that $\mathbf{b}$

points to the right. (Rotation actually changes the vectors, of course, but does not change the scalar projection.) Now, the cosine of the angle θ between $\mathbf{a}$ and $\mathbf{b}$ is the scalar projection of $\hat{\mathbf{a}}$ onto $\mathbf{b}$, from the definition of the cosine function. By similar triangles, $\left|\text{proj}_{\mathbf{b}}\,\mathbf{a}\right|$ is $\|\mathbf{a}\|$ times this. Therefore we find:

$$\left|\text{proj}_{\mathbf{b}}\,\mathbf{a}\right| = \|\mathbf{a}\|\cos\theta = \|\mathbf{a}\|\frac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|} = \frac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{b}\|}$$

and *vice versa*. We could also write this as

$$\left|\text{proj}_{\mathbf{b}}\,\mathbf{a}\right| = \mathbf{a}\cdot\hat{\mathbf{b}}$$

because the dot product associates with scalar multiplication. To compute a vector projection, we notice that we need a vector with the directed length $\left|\text{proj}_{\mathbf{b}}\,\mathbf{a}\right|$ along $\mathbf{b}$. This is accomplished by scaling the unit vector $\hat{\mathbf{b}}$ by the value of the scalar projection:

$$\text{proj}_{\mathbf{b}}\,\mathbf{a} = \frac{\mathbf{a}\cdot\mathbf{b}}{\|\mathbf{b}\|}\hat{\mathbf{b}}$$

We could also write this as

$$\text{proj}_{\mathbf{b}}\,\mathbf{a} = (\mathbf{a}\cdot\hat{\mathbf{b}})\hat{\mathbf{b}}$$

Two notes: first, projecting onto the zero vector is meaningless since it has no direction, and second, neither scalar nor vector projection is commutative.

Determinant   [edit]

The *determinant* of two vectors $\mathbf{a}$ and $\mathbf{b}$, denoted $\det(\mathbf{a},\mathbf{b})$, is the determinant of the matrix having $\mathbf{a}$ and $\mathbf{b}$ as rows (or, equivalently, columns), that is:

$$\begin{bmatrix} a_x & a_y \\ b_x & b_y \end{bmatrix}$$

and is therefore equal to $a_x b_y - a_y b_x$. The size-2 determinant is often mistakenly referred to as the *cross product*. (This usage is nearly universal in casual programming parlance and less reputable websites.) This is because there exists a product of two three-dimensional vectors known as the cross product, and the cross-product of two vectors lying in the xy-plane is a vector pointing along the z-axis with z-component equal to the determinant of the two plane vectors. However, no background with three-dimensional vectors, cross-products, or matrices is required for this section. In two dimensions, all that is needed is the formula: $\det(\mathbf{a}, \mathbf{b}) = a_x b_y - a_y b_x$. However, for the sake of convenience, we shall denote the size-2 determinant by the same symbol as the cross product, the cross: hence $\mathbf{a} \times \mathbf{b} = a_x b_y - a_y b_x$. (This also avoids confusion with the larger matrices in advanced linear algebra and higher dimensions.) It is anticommutative, meaning that $\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$. It satisfies the properties $\alpha(\mathbf{a} \times \mathbf{b}) = (\alpha\mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (\alpha\mathbf{b})$, and with addition it is both left-distributive, so that $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$, and right-distributive, so that $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$. There are two primary uses for the size-2 determinant in two-dimensional computational geometry.

The determinant is useful in several algorithms for computing the convex hull because it allows us to determine the *directed angle* from one vector to another. (Note that the dot product allows us to determine the undirected angle; it makes sense because the dot product commutes whereas the size-2 determinant anticommutes.) The following relation holds:

$$\mathbf{a} \times \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \sin\theta$$

where θ is the *directed* angle from $\mathbf{a}$ to $\mathbf{b}$. That is, the counterclockwise angle from $\mathbf{a}$ to $\mathbf{b}$ when they are both anchored at the same point. Rearranging allows us to compute the angle directly:

$$\theta = \sin^{-1} \frac{\mathbf{a} \times \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$$

(Note that the sine function is odd, so that when $\mathbf{a}$ and $\mathbf{b}$ are interchanged in the above, the right side is negated, and so is the left.) Often, however, we do not need the directed angle itself. In the monotone chains algorithm for convex hull computation, for example, all we need to know is whether there is a clockwise or a counterclockwise turn from $\mathbf{a}$ to $\mathbf{b}$. (That is, if one wishes to rotate $\mathbf{a}$ through no more than 180 degrees either clockwise or counterclockwise so that its direction coincides with that of $\mathbf{b}$, which direction should one choose?) It turns out that if $\mathbf{a} \times \mathbf{b}$ is positive, the answer is counterclockwise; if it is negative, clockwise, and if it is zero, the vectors are either parallel or antiparallel. (Indeed, the closer θ is to a right angle, the larger the cross product in magnitude; in some sense then it measures "how far away" the vectors are from each other.)

The determinant is also useful in computing areas of parallelograms and triangles. If the two vectors $\mathbf{a}$ and $\mathbf{b}$ are placed tail-to-tail, a parallelogram can be formed using these vectors as adjacent sides; its area is $\left| \mathbf{a} \times \mathbf{b} \right|$. A triangle can also be formed using these vectors as adjacent sides; its area is half that of the parallelogram, or $\frac{1}{2} |\mathbf{a} \times \mathbf{b}|$. (This echoes the formula $A = \frac{1}{2} ab \sin \theta$ taught in high-school math classes.) In fact, the area of a general quadrilateral $ABCD$ can be computed with a determinant as well; it equals $\frac{1}{2} |(C - A) \times (D - B)|$.

Circles   [edit]

Introduction to circles   [edit]

A circle is *the locus of points in the plane equidistant from a given point*. That is, we choose some point O , the *centre*, and some distance $r > 0$, the *radius*, and the circle consists exactly of those points whose Euclidean distance from O is exactly r. When storing a circle in memory, we store merely the centre and the radius.

Equation of a circle [edit]

Suppose the circle has centre $O(h, k)$ and radius r. Then, from the definition, we know that any point (x, y) on the circle must satisfy $\text{dist}((x, y), (h, k)) = r$. This means $\sqrt{(x - h)^2 + (y - k)^2} = r$, or $(x - h)^2 + (y - k)^2 = r^2$.

Inside, outside, or on the circle [edit]

The equation of a circle is a sufficient and necessary condition for a point to be on the circle. If it is not on the circle, it must be either inside the circle or outside the circle. It will be inside the circle when its distance from the centre is less than r, or $(x - h)^2 + (y - k)^2 < r^2$, and similarly it will be outside the circle when $(x - h)^2 + (y - k)^2 > r^2$.

Intersection of a circle with a line [edit]

To determine points of intersection of a circle with another figure (it might also be a circle), solve the simultaneous equations obtained in x and y. For example, given a circle centered at (2,3) with radius 2, and the line $x + y - 4 = 0$, we would solve the simultaneous equations $(x - 2)^2 + (y - 3)^2 = 2^2$ and $x + y - 4 = 0$. If there are multiple solutions, each is a different point of intersection; if there are no solutions then the two figures do not intersect. Here then are general results. Note that if you do not check the "no intersection" condition beforehand

and plunge straight into the quadratic formula (after reducing the two simultaneous equations to one equation), you will try to extract the square root of a negative number, which will crash some languages (such as Pascal).

No points of intersection   [edit]

When the closest distance from the centre of the circle to the line is greater than the radius of the circle, the circle and line do not intersect. (The formula for the distance from a point to a line can be found in the Lines section of this article.)

One point of intersection   [edit]

When the closest distance from the centre of the circle to the line is exactly the circle's radius, the line is tangent to the circle. One way of finding this point of tangency, the single point of intersection, is to drop a perpendicular from the centre of the circle to the line. (The technique for doing so is found in the Lines section.) This will, of course, yield the same answer as solving the simultaneous equations.

Two points of intersection   [edit]

When the closest distance from the centre of the circle to the line is less than the circle's radius, the line intersects the circle twice.

To find the two points of intersection, first drop a perpendicular from the centre of the circle to the line. Call the foot of the perpendicular P and denote the distance $|OP|$ by d. Now, let the two points of intersection be Q_1 and Q_2. Then, the triangles $\triangle OPQ_1$ and $\triangle OPQ_2$ are right-angled, and hence we see that $|OP|^2 + |Q_1P|^2 = |OQ_1|^2 = r^2$ and $|OP|^2 + |Q_2P|^2 = |OQ_2|^2 = r^2$. So we compute the value $h = \sqrt{r^2 - |OP|^2}$; now,

if we move h units in either direction along the line from P, we will reach either Q_1 or Q_2. (The technique to do so is given in section [Translation along a line](#).)

Intersection of a circle with a circle [[edit](#)]

Finding the points of intersection of two circles follows the same basic idea as the circle-line intersection. Here's how to determine the nature of the intersection beforehand, to avoid accidentally trying to take the square root of a negative number:

No points of intersection [[edit](#)]

When the distance between the centres of the circles is less than the difference between their radii, the circle with smaller radius will be contained completely within the circle of larger radius. When the distance between the centres of the circles is greater than the sum of their radii, neither circle will be inside the other, but still the two will not intersect.

One point of intersection [[edit](#)]

Let the two circles have centres O_1 and O_2 and radii r_1 and r_2.

- When the distance between the centres of the circles is exactly the difference between their radii, the two circles will be internally tangent. Suppose that the first circle is larger. Then the point of intersection lies on the line between O_1 and O_2. The intersection point and O_1 are on opposite sides of O_2, and its distances from O_1 and O_2 are r_1 and r_2, respectively. This gives $P = \dfrac{O_2 r_1 - O_1 r_2}{r_1 - r_2}$. (It turns out that if we swap the 1's and 2's in this formula, we get the same value; hence we can use this formula without regard to which circle is labelled 1 and which one is 2.)

- When the distance between the centres of the circles is exactly the sum of their radii, they will be externally tangent; they will have one point of intersection located between O_1 and O_2; it will be r_1 units away from the former and r_2 units away from the latter, giving the formula

$$P = \frac{O_2 r_1 + O_1 r_2}{r_1 + r_2}.$$

Two points of intersection   [edit]

In all other cases, there will be two points of intersection.

To find them, it will be helpful to draw a diagram. Denote the centres of the circles O_1, O_2, their radii r_1, r_2, and the two points of intersection P_1, P_2. First draw the *radical axis*, the line that passes through P_1 and P_2. This passes through the midpoint M between P_1 and P_2, and here it intersects perpendicularly the line between O_1 and O_2. Denote $|O_1 M|$ and $|O_2 M|$ by d_1, d_2, respectively. Clearly

$$d_1 + d_2 = |O_1 O_2| \text{ . (1)}$$

Also, from the Pythagorean theorem,

$$|MP_1|^2 = |O_1 P_1|^2 - |O_1 M|^2 = |O_2 P_1|^2 - |O_2 M|^2 \text{, or } r_1^2 - d_1^2 = r_2^2 - d_2^2.$$

Rearranging this gives

$$d_1^2 - d_2^2 = r_1^2 - r_2^2. \text{ (2)}$$

Now we solve for d_1 and d_2. To do this, we factor the left side of (2), giving

$$(d_1 - d_2)(d_1 + d_2) = r_1^2 - r_2^2, \text{ which allows us to divide through by (1), thus obtaining}$$

$$d_1 - d_2 = \frac{r_1^2 - r_2^2}{|O_1 O_2|} \text{ . (3)}$$

Equations (1) and (3) are linear and may be used to solve for d_1, d_2 in a straightforward fashion.

Now we can compute the coordinates of M as $M = \dfrac{O_2 d_1 + O_1 d_2}{d_1 + d_2}$, using the argument from the previous section. Then construct the line perpendicular to $O_1 O_2$ that passes through M (the radical axis). Finally, we can find the distance $|MP_1| = |MP_2| = \sqrt{r_1^2 - d_1^2} = \sqrt{r_2^2 - d_2^2}$, and move this distance in either direction along the radical axis (see Translation along a line) to obtain P_1 and P_2.

It should be noted that two circles that intersect at two points cut each other into two arcs each: one arc that lies within the other circle, and one that lies outside it. It is not immediately obvious which is which. However, you should be able to convince yourself from drawing a diagram that if $(P_1 - O_1) \times (O_2 - O_1) > 0$, then the arc of the first circle that is inside the second circle is counterclockwise from P_1 to P_2, but if that determinant is negative, then the inside arc is counterclockwise from P_2 to P_1. (For an example of where this is important, see Detectors 🔗.)

Tangents   [edit]

A line is said to be *tangent* to a circle if it intersects the circle at exactly one point.

The tangent has an important property: if a circle with centre O has a tangent line l that touches it at P, then $l \perp OP$. The proof of this fact is simple: we assume that the two lines are not perpendicular, then drop a perpendicular from O to l at Q. We observe that $\triangle OPQ$ is right-angled at Q, and that therefore $|OP| > |OQ|$. But this implies that Q is inside the circle, a contradiction.

To actually find the tangent line to one circle at a given point on its circumference is easy; it is simply the line passing through that point that is perpendicular to the radius.

From point to circle [edit]

Sometimes we may also be given a circle with centre and radius (O, r) plus a point P and asked to find a tangent to the circle that passes through the point. There may be either zero, one, or two such tangents.

Point is inside circle [edit]

When the point is inside the circle, there are no tangents to the circle through that point, because any line through the point must intersect the circle twice.

Point is on circle [edit]

When the point is on the circumference of the circle, the line passing through P that is perpendicular to OP is the unique tangent to the circle passing through that point.

Point is outside circle [edit]

When the point is outside the circle, there are two tangents to the circle passing through the point. To construct one such tangent, we use a trick called *inversion*. We let one of the points of tangency be denoted Q. Now $\triangle OQP$ is right-angled at Q. If we now drop a perpendicular from Q to OP at R, we obtain triangle $\triangle OQR$ which is similar to $\triangle OPQ$. This means that $\dfrac{|OP|}{|OQ|} = \dfrac{|OQ|}{|OR|}$, so $|OR| = \dfrac{r^2}{|OP|}$. Note that R is *inside* the circle, unlike P which is *outside* (hence the name *inversion*.) This allows us to locate point R on the line segment OP. Then, we see that the line segment RQ is perpendicular to OP and has length $\sqrt{r^2 - |OR|^2}$ from the Pythagorean theorem in $\triangle OQR$. This gives us enough information

to compute the coordinates of Q. There will be two possibilities, since we can draw two rays through R that are perpendicular to OP.

Circle-circle tangents   [edit]

A pair of circles may have either zero, one, two, three, or four mutual tangents.

Zero mutual tangents   [edit]

If one circle lies within the other (that is, the difference between their radii is greater than the distance between their centres), then they have no mutual tangents, since every tangent to the outer circle lies outside it, and hence cannot touch the inner circle.

One mutual tangent   [edit]

Two circles will have one mutual tangent if they are internally tangent (that is, the difference between their radii equals the distance between their centres). This tangent can easily be found after constructing the point P of internal tangency (see section One point of intersection above).

Two mutual tangents   [edit]

A pair of circles that intersect at two points has two mutual tangents, one on either "side". We will consider the problem of how to determine one of them.

Assume the circles have centres O_1, O_2 and radii r_1, r_2. Assume further that they have some mutual tangent l and that l is tangent to the circles at P_1, P_2, respectively. We see that $O_1P_1 \perp l$ and $O_2P_2 \perp l$. Now, assume without loss of generality that $r_2 \geq r_1$. Then drop the perpendicular from O_1 to O_2P_2 at Q, and obtain $|O_2Q| = r_2 - r_1$. Then

$$\theta = \angle O_1O_2Q = \cos^{-1}\frac{r_2 - r_1}{|O_1O_2|}.$$ Also, we can solve for $h = |O_1Q|$ using the

Pythagorean theorem, obtaining $h = \sqrt{|O_1O_2|^2 - (r_2 - r_1)^2}$. Notice that because $O_1QP_2P_1$ is a rectangle, $h = |P_1P_2|$.

There are two possibilities for the location of P_2. One of them is obtained by rotating the ray $O_1 - O_2$ clockwise through angle θ, the other obtained by rotating it counterclockwise, and in either case determining where the ray cuts the second circle. (So it doesn't actually matter which circle is bigger; the formula $\cos^{-1}\dfrac{r_2 - r_1}{|O_1O_2|}$ will give the same result up to sign.)

We can then argue that the coordinates of one of the possible P_2's is given by $O_2 + r_2[\cos(\mathrm{atan2}(O_1 - O_2) + \theta), \sin(\mathrm{atan2}(O_1 - O_2) + \theta)]$, because $\mathrm{atan2}(O_1 - O_2)$ represents the directed angle from the positive x-axis to the ray from O_2 to O_1, and we are adding onto this directed angle the value θ to get our final position. The other possible P_2 will be given by subtracting θ instead of adding.

This formula as written is slow, since it involves trigonometric functions, which have a higher invisible constant factor than algebraic operations and square roots. Therefore, we will rewrite it in terms of only algebraic operations. Define $\Delta x = O_{1_x} - O_{2_x}$ and Δy analogously.

Verify the identities $\cos\mathrm{atan2}\dfrac{y}{x} = \dfrac{x}{x^2 + y^2}$ and $\sin\mathrm{atan2}\dfrac{y}{x} = \dfrac{y}{x^2 + y^2}$. (These are actually obvious from the definition of `atan2`). Then:

$$\cos(\mathrm{atan2}(O_1 - O_2) + \theta)$$
$$= \cos\mathrm{atan2}(O_1 - O_2)\cos\theta - \sin\mathrm{atan2}(O_1 - O_2)\sin\theta$$
$$= \frac{\Delta x}{\Delta x^2 + \Delta y^2}\frac{r_2 - r_1}{|O_1O_2|} - \frac{\Delta y}{\Delta x^2 + \Delta y^2}\frac{h}{|O_1O_2|}$$

$$= \frac{(r_2 - r_1)\Delta x - h\Delta y}{|O_1 O_2|^2}$$

And likewise

$$\sin(\text{atan2}(O_1 - O_2) + \theta)$$
$$= \sin \text{atan2}(O_1 - O_2) \cos\theta + \cos \text{atan2}(O_1 - O_2) \sin\theta$$
$$= \frac{\Delta y}{\Delta x^2 + \Delta y^2} \frac{r_2 - r_1}{|O_1 O_2|} + \frac{\Delta x}{\Delta x^2 + \Delta y^2} \frac{h}{|O_1 O_2|}$$
$$= \frac{h\Delta x + (r_2 - r_1)\Delta y}{|O_1 O_2|^2}$$

After locating P_2, we just construct the line passing through P_2 that is perpendicular to $O_2 P_2$ and this will also be tangent to the first circle at P_1.

Three mutual tangents   [edit]

The case of three mutual tangents occurs when the two circles are externally tangent that is, when the sum of their radii equals the distance between their centres. In this case two of the three tangents (the "outer" ones) can be found using the formula in the previous section, and the third, which passes "between" the circles, may easily be found by first finding the point of mutual tangency (see section One point of intersection above).

Four mutual tangents   [edit]

If two circles do not intersect or overlap at all, that is, the distance between their centres is greater than the sum of their radii, they will have four mutual tangents. Two of the mutual tangents can be found using the formulas two sections above. These are the two "outer" tangents, with the property

that both circles lie on the same side of an outer tangent. There are also two "inner" tangents that "separate" the two circles from each other.

The procedure for computing these "inner" tangents is similar to the procedure for computing the "outer" tangents. We begin by letting l be one of these tangents, that touches the first circle at P_1 and the second at P_2. Let the point of intersection of l with O_1O_2 be denoted Q. Then, right-angled $\triangle O_1 P_1 Q$ is similar to right-angled $\triangle O_2 P_2 Q$. Let $d_1 = |O_1Q|$ and $d_2 = |O_2Q|$. Then, by similarity, we see that $\dfrac{d_1}{r_1} = \dfrac{d_2}{r_2}$. Along with the equation

$d_1 + d_2 = |O_1O_2|$, we can solve: $d_1 = \dfrac{r_1}{r_1 + r_2}|O_1O_2|, d_2 = \dfrac{r_2}{r_1 + r_2}|O_1O_2|$.

Then, letting $\theta = \angle P_2 O_2 Q$, we have $\cos\theta = \dfrac{r_2}{d_2} = \dfrac{r_1 + r_2}{|O_1O_2|}$. Let $k = |P_1 P_2|$, and

see that $k = |QP_1| + |QP_2| = \sqrt{d_1^2 - r_1^2} + \sqrt{d_2^2 - r_2^2}$, which, after some messy

algebra, will turn out to be $\sqrt{|O_1O_2|^2 - (r_1 + r_2)^2}$. Also observe that

$\sin\theta = \dfrac{|P_2Q|}{|O_2Q|} = \dfrac{|P_1Q|}{|O_1Q|} = \dfrac{|P_1Q| + |P_2Q|}{|O_1Q| + |O_2Q|} = \dfrac{k}{|O_1O_2|}$.

Moving along, we have

$P_2 = O_2 + r_2\big[\cos(\text{atan2}(O_1 - O_2) + \theta), \sin(\text{atan2}(O_1 - O_2) + \theta)\big]$ as before.
Defining $\Delta x = O_{1_x} - O_{2_x}$ and Δy analogously, we obtain:

$$\cos(\text{atan2}(O_1 - O_2) + \theta)$$
$$= \cos\text{atan2}(O_1 - O_2)\cos\theta - \sin\text{atan2}(O_1 - O_2)\sin\theta$$

$$= \frac{\Delta x}{\Delta x^2 + \Delta y^2} \frac{r_1 + r_2}{|O_1 O_2|} - \frac{\Delta y}{\Delta x^2 + \Delta y^2} \frac{k}{|O_1 O_2|}$$

$$= \frac{(r_1 + r_2)\Delta x - k\Delta y}{|O_1 O_2|^2}$$

and likewise

$$\sin(\text{atan2}(O_1 - O_2) + \theta)$$
$$= \sin \text{atan2}(O_1 - O_2) \cos \theta + \cos \text{atan2}(O_1 - O_2) \sin \theta$$
$$= \frac{\Delta y}{\Delta x^2 + \Delta y^2} \frac{r_1 + r_2}{|O_1 O_2|} + \frac{\Delta x}{\Delta x^2 + \Delta y^2} \frac{k}{|O_1 O_2|}$$
$$= \frac{k\Delta x + (r_1 + r_2)\Delta y}{|O_1 O_2|^2}$$

Careful! The value of h in this case is not the same as the value of h for the outer tangents.

Summary and discussion   [edit]

We will summarize the results below. (Note that we have exchanged the two circles.) We define

$h = \sqrt{|O_1 O_2|^2 - (r_1 - r_2)^2}$ and $k = \sqrt{|O_1 O_2|^2 - (r_1 + r_2)^2}$, and then the points

of tangency of the two outer tangents with the first circle are given by:

$$P_1 = O_1 + \frac{r_1}{|O_1 O_2|^2}[(r_1 - r_2)\Delta x - h\Delta y, h\Delta x + (r_1 - r_2)\Delta y]$$

$$P_2 = O_1 + \frac{r_1}{|O_1 O_2|^2}[(r_1 - r_2)\Delta x + h\Delta y, -h\Delta x + (r_1 - r_2)\Delta y]$$

and likewise for the two inner tangents:

$$P_3 = O_1 + \frac{r_1}{|O_1 O_2|^2}[(r_1 + r_2)\Delta x - k\Delta y, k\Delta x + (r_1 + r_2)\Delta y]$$

$$P_4 = O_1 + \frac{r_1}{|O_1 O_2|^2}[(r_1 + r_2)\Delta x + k\Delta y, -k\Delta x + (r_1 + r_2)\Delta y]$$

The first pair of formulas will be invalid when $|r_1 - r_2| > |O_1 O_2|$, because h will be undefined. Obviously, in this case, $r_1 + r_2 > |O_1 O_2|$ also, so k will also be undefined, and the second pair of formulas will be invalid.

When $|r_1 - r_2| = |O_1 O_2|$, we will obtain $h = 0$, so P_1 and P_2 will coincide. The second pair of formulas will still be invalid, so we will obtain a total of one point of tangency.

When $|r_1 - r_2| < |O_1 O_2| < r_1 + r_2$, we will obtain $h > 0$, so P_1 and P_2 will be distinct, but k will still be undefined, so we obtain a total of two points of tangency.

When $|O_1 O_2| = r_1 + r_2$, we will obtain $h > 0, k = 0$ giving two distinct points P_1, P_2, but $P_3 = P_4$, a total of three points of tangency.

Finally, when $|O_1 O_2| > r_1 + r_2$, then $h, k \geq 0$ and all four points P_1, P_2, P_3, P_4 will be distinct, as expected.

Categories: Algorithms | Geometry | Pages needing diagrams