Behind the Sat-Nav

Understanding the mathematics and algorithms behind an everyday item

"SatNav" is a colloquial word for Satellite Navigation. It is a decades old technology first used by a US military program called *Transit.* Since then, SatNav technology has become increasingly widespread – currently there are four global satellite networks:

- the United States's Global Positioning System (GPS),
- Russia's Global Navigation Satellite System (GLONASS),
- China's BeiDou Navigation system (BDS)
- European Space Agency's Galileo

In today's world SatNavs are almost universal, gone are the days of buying maps to work out a route from A to B, before cursing and swearing because you took a wrong turn, and with the advent of smartphones such technology is made even more accessible with apps such as *Apple Maps* or *Waze.*

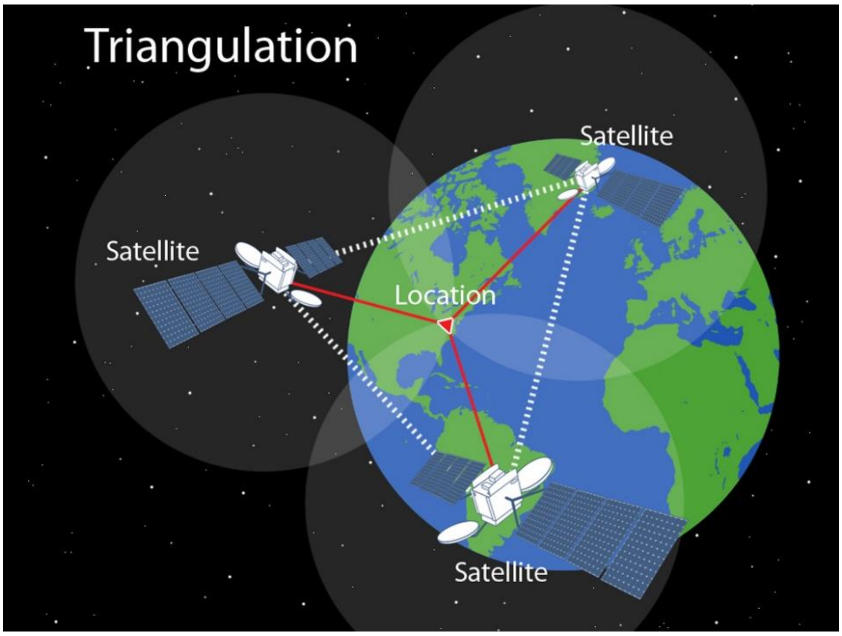
Credit: pcjumbo

At this point it is key to remember that building a satellite network is only half the battle. If you wish to make the next *Google Maps* you also need a system to tell you the directions to get to your destination. This relies on a completely different technology involving graphs – which will be covered later.

Credit: New Scientist

Determining locations

Satellites use a process called “trilateration” to determine your position on the earth. The satellites are constantly emitting data about their own location, as well as an accurate time of the measurement of that data – found with an atomic clock.



Credit: National Geographic Society

Your SatNav receives a radio broadcast of the position of three separate satellites as well as the time these measurements were simultaneously taken. Since radio waves travel at the constant speed of light, your SatNav can use the time elated between the signal being sent and the signal being received to determine its distance to each satellite respectively. However, since oftentimes 3 distinct spheres can intersect at 2 points a fourth satellite may be needed.

Suppose that our 3 satellites have positions $(X_1, Y_1, Z_1), (X_2, Y_2, Z_2)$ and (X_3, Y_3, Z_3) with radii R_1, R_2 and R_3 respectively. Suppose also that our SatNav is in a car with coordinates (x_c, y_c, z_c). Then we can write the following:

$$(x_c - X_1)^2 + (y_c - Y_1)^2 + (z_c - Z_1)^2 = R_1^2$$

$$(x_c - X_2)^2 + (y_c - Y_2)^2 + (z_c - Z_2)^2 = R_2^2$$

$$(x_c - X_3)^2 + (y_c - Y_3)^2 + (z_c - Z_3)^2 = R_3^2$$

Expanding these:

$$(1)\ x_c^2 - 2x_cX_1 + X_1^2 + y_c^2 - 2y_cY_1 + Y_1^2 + z_c^2 - 2z_cZ_1 + Z_1^2 = R_1^2$$

$$(2)\ x_c^2 - 2x_cX_2 + X_2^2 + y_c^2 - 2y_cY_2 + Y_2^2 + z_c^2 - 2z_cZ_2 + Z_2^2 = R_2^2$$

$$(3)\ x_c^2 - 2x_cX_3 + X_3^2 + y_c^2 - 2y_cY_3 + Y_3^2 + z_c^2 - 2z_cZ_3 + Z_3^2 = R_3^2$$

We can then subtract one equation from the another to give us the equation for the plane which contains the points on each sphere that are shared.

$(1) - (2)$ =

$$2x_c(X_2 - X_1) + X_1^2 - X_2^2 + 2y_c(Y_2 - Y_1) + Y_1^2 - Y_2^2 + 2z_c(Z_2 - Z_1) + Z_1^2 - Z_2^2 = R_1^2 - R_2^2$$

Then let $\alpha_1 = 2(X_2 - X_1), \beta_1 = 2(Y_2 - Y_1), \gamma_1 = 2(Z_2 - Z_1)$

Also define $C_1 = R_1^2 - R_2^2 - X_1^2 + X_2^2 - Y_1^2 + Y_2^2 - Z_1^2 + Z_2^2$

Then we can write

$(1) - (2)$ = $\alpha_1 x_c + \beta_1 y_c + \gamma_1 z_c = C_1$

Doing a similar substitution for $(1) - (3)$ and $(2) - (3)$ we have the 3 planes:

$$\alpha_1 x_c + \beta_1 y_c + \gamma_1 z_c = C_1$$

$$\alpha_2 x_c + \beta_2 y_c + \gamma_2 z_c = C_2$$

$$\alpha_3 x_c + \beta_3 y_c + \gamma_3 z_c = C_3$$

Which can be solved to find an exact value for the SatNav's location.

Time Dilation and Satellites

Time Dilation is a phenomenon first proposed by Albert Einstein in 1905, he proposed that time passes differently for two observers in motion or in different gravitational fields.

Einstein's theory of *General relativity* explained the effect large masses have on how time is perceived, whilst the theory of *Special relativity* explained the relationship between high speeds and the perception of time.

This has implications for GPS systems, firstly, an object on earth is in a greater gravitational field than a satellite in orbit, and the satellite is moving much faster relative to a SatNav on the earth. This means that there are two sources of error that need to be accounted for.

We can compute the error caused by Special relativity by using the *Lorentz factor* to determine the exact difference in time. We have:

$$t = t_0 \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Where t is the time on earth, t_0 is the time as measured on the satellite, v is the speed of the satellite and c is the speed of light.

GPS satellites travel at an altitude of ~20,200km which means they have a speed of about ~3871m/s. Even though the satellite is travelling this quickly the Lorentz factor is still very small, at just 8.32×10^{-11} greater than 1.

We have:

$$\frac{\Delta t}{t} = \frac{t - t_0}{t} = 1 - \frac{t_0}{t}$$

For simplicity, put $t = 1s$. Then we have $\frac{\Delta t}{t} = 8.32 \times 10^{-11}$. This means that the time dilation per second is 8.32×10^{-11} seconds. Though this may not seem like much, given that there are $604{,}800$ seconds in a week. That means the total time dilation from the effects of speed alone is about 50 microseconds after a week.

This is still a small number; however, it is a significant amount in the context of navigation. In 50 microseconds, light travels 15km. A Satnav that had a margin of error of 15km wouldn't be much good.

We also need to measure the error caused by General relativity, that is, the difference in perception of time as a result of the large gravitational field of the earth. Since there is a significant effect on both the satellite and the earth's surface it is best to do this calculation with the frame of reference of an empty point deep in space having no time dilation.

We can use the formula:

$$t = t_0 \frac{1}{\sqrt{1 - \frac{2GM}{Rc^2}}}$$

Where t is the time on earth, t_0 is the time on the satellite, R is the radius of the earth/orbit, G is the universal gravitation constant, M is the mass of the earth and c is the speed of light.

Then we can manipulate to find $\frac{\Delta t}{t}$

$$\frac{\Delta t}{t} = \frac{t - t_0}{t} = 1 - \frac{t_0}{t}$$

$$t = t_0 \frac{1}{\sqrt{1 - \frac{2GM}{Rc^2}}} \Rightarrow \sqrt{1 - \frac{2GM}{Rc^2}} = \frac{t_0}{t}$$

$$\Rightarrow \frac{\Delta t}{t} = 1 - \sqrt{1 - \frac{2GM}{Rc^2}}$$

On the earth surface the time dilation is ~3.47×10^{-10}

On the satellite the time dilation is ~ 8.33×10^{-11}

So the relative time dilation between the satellite and the SatNav is 2.637×10^{-10}seconds per second.

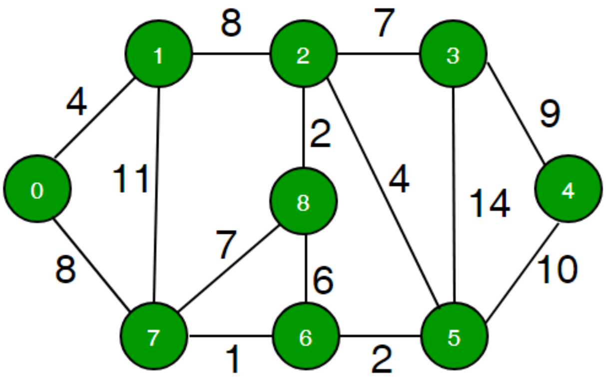
To put this into context, after one week (604,800 seconds) the total time dilation is 1.6×10^{-4} seconds.

Pathfinding

In the modern day, SatNavs use a combination of both pathfinding algorithms and proprietary data such as traffic, weather and routes commonly taken by other users. But at a basic level, pathfinding algorithms that have been known for decades are still used on a day-to-day basis.

Maps are often abstracted to become graphs, with junctions and crossroads being nodes whilst roads are represented as vertices. The fastest known pathfinding algorithm is known as the A* search algorithm, which itself is an optimised version of Dijkstra's algorithm.

An example of a graph: each vertex has being given a "weight"



Credit: Geeks for Geeks

The first step is to compare all of the edges that connect to the starting node and investigate which one is the shortest. Label each connecting node with the distance it takes to get to them as well as the current shortest known path, then "visit" the node with the shortest distance. Once again, label each connecting node with the distance to them and the current best-known path. If a node has already been labelled, you "update" the label with the shorter path. This process is repeated until all nodes are visited, giving you the shortest path to each node

Dijkstra's algorithm is widely used and learned because of it's simplicity. Also the fact that visiting a node guarantees the shortest path has been found makes it particularly relevant in the context of SatNavs as once the algorithm has been executed the shortest paths to *all nodes* are found and stored. So a user could execute the algorithm in their local area and have routes from the home to nearby places stored.

The A* search algorithm is an optimisation of Dijkstra's algorithm that uses "heuristics" or rule-based systems to lessen the time taken to find the shortest path. Within the context of SatNavs, an example could be the Manhattan distance (distance in x + distance in y) or simple Euclidean distance. The heuristic function is used to guide the search of the graph, so certain nodes are visited before others – when they would not be in standard Dijkstra's algorithm.

To conclude, SatNavs in everyday operation are incredibly mathematical – even ignoring the orbital mechanics and advanced physics required to send a satellite into orbit. It is worth mentioning that technical discussions of things like SatNavs can help to get people interested in mathematics and science – and this topic could be a useful part of mathematics outreach in the future.