# Geodesics on a Cone[1]

A geodesic on a surface is a curve, connecting two given points, such that any nearby curve with the same endpoints is longer. For instance, geodesics in $\mathbb{R}^n$ are straight lines, on the sphere they are great circles. In the latter case, given two points $A$, $B$ draw the great circle centered at the sphere's centre and passing through $A$ and $B$. Both pieces of the circumference are geodesics, although normally one will be shorter than the other. Both corresponds to the local minima of the curve length functional.

Let us now consider geodesics on the surface of the cone. The number of distinct geodesics connecting two points will always turn out to be finite, and will depend essentially only on the angle of the cone, plus there will be a few special cases, corresponding to specifically chosen boundary conditions. Distinct geodesics connecting the two points will differ by how many times they wrap around the cone, and the shortest one will obviously have the smallest angle change.

Before starting, observe that if one cuts the cone along its edge, the cone unwraps into a sector of the Euclidean plane, and the geodesics on the cone must yield straight line segments in the sector. The role of geometry, however is striking. For each pair of points $A, B$ there will be a finite number of geodesics, connecting them, each one being characterised by its *winding number*, that is the number of times it wraps around the cone's axis. A geodesic with a given winding number $m$ will be the shortest line connecting $A$ and $B$ in the class of limes connecting $A$ and $B$ and wrapping $m$ times around the cone.

Since the rays bounding the sector are glued together on the cone, geodesics on the cone will generally become broken lines after it has been cut to yield the plane sector. The geodesic with wrapping number zero will be simply the straight line connecting the points $A$ and $B$ after the cone has been unwrapped into the sector. Geodesics with higher winding numbers will be broken lines that after reaching the edge of the sector continue as straight lines off the other edge, so that the corresponding segments of the broken line form equal angles with the sector's two edges.

## 1.1 Coordinates.

We consider a right circular cone (so that the cone axis is perpendicular to the circular base) of indefinite height. Take the apex of the cone to be at the origin and the axis to be along the positive $z$-axis. Let $\alpha$ denote the half-angle of the cone, ie the angle between the axis and a ray along the cone edge. Then, at a point $(x, y, z)$ on the cone,

$$\rho := (x^2 + y^2)^{1/2} = \tan\alpha\, z = kz, \tag{1}$$

where $k = \tan\alpha$. Introducing polar coordinates $(\rho, \phi)$ for $x$ and $y$, we have that

$$\begin{aligned} x &= \rho\cos\phi = kz\cos\phi, \\ y &= \rho\sin\phi = kz\sin\phi. \end{aligned} \tag{2}$$

Thus, points on the cone are parameterised by $z$ and $\phi$.

Observe that in polar coordinates, the infinitesimal change $\rho \to \rho + d\rho$, $\phi \to \phi + d\phi$ implies the increments

$$\begin{aligned} dx &= d\rho\cos\phi - \rho\sin\phi d\phi \\ dy &= d\rho\sin\phi - \rho\cos\phi d\phi, \end{aligned} \tag{3}$$

and therefore, using $\sin^2\phi + \cos^2\phi = 1$,

$$dx^2 + dy^2 = d\rho^2 + \rho^2 d\phi^2. \tag{4}$$

Since the cone has symmetry with respect to rotations around the $z$-axis it is natural to use cylindrical coordinates $(\rho, \phi, z)$ to describe it, which are just polar coordinates in the $(x, y)$ plane with the $z$-coordinate added.

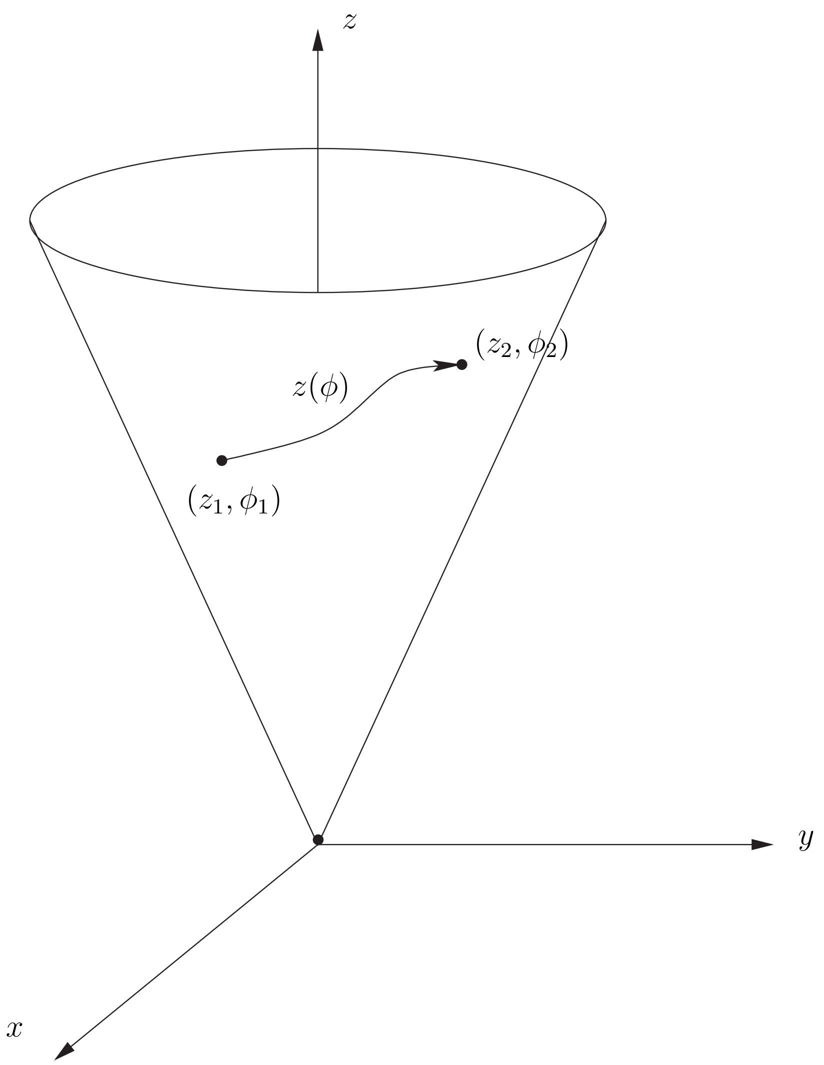
[1] Based on J. Robbins' notes.

### 1.2 Length of curve

We parameterise a curve on the cone by specifying $z$ as a function of $\phi$, ie $z = z(\phi)$. This leaves out the rays $\phi = const.$, which are only possible with specific boundary conditions $\phi_1 = \phi_2$ at the curve's endpoints, are obviously geodesics. In addition, using the angle $\phi$ as a coordinate does not allow us to deal with the apex, as for $x = y = 0$ the angle $\phi$ is simply not defined.

The curve is to have fixed endpoints. Let $(z_1, \phi_1)$ denote the coordinates of the initial point, and $(z_2, \phi_2)$ the coordinates of the final point. Then $z(\phi)$ must satisfy the boundary conditions

$$z(\phi_1) = z_1, \quad z(\phi_2) = z_2. \tag{5}$$




The cartesian coordinates along the curve are given by

$$\begin{aligned} x &= x(\phi) = kz(\phi)\cos\phi, \\ y &= y(\phi) = kz(\phi)\sin\phi, \\ z &= z(\phi), \end{aligned} \tag{6}$$

while in the cylindrical coordinates we have

$$\begin{aligned}\rho &= kz(\phi),\\ \phi &= \phi,\\ z &= z(\phi).\end{aligned} \tag{7}$$

The length $ds$ of the infinitesimal segment between $\phi$ and $\phi + d\phi$ is given by

$$ds^2 = dx^2 + dy^2 + dz^2 = d\rho^2 + \rho^2 d\phi^2 + dz^2, \tag{8}$$

now using $dz = z' d\phi$ and $\rho = kz$, we have

$$ds = \left((1+k^2)z'^2 + k^2 z^2\right)^{1/2} d\phi. \tag{9}$$

Then the length of the curve $z(\phi)$ is given by

$$L = \int ds = \int_{\phi_1}^{\phi_2} \left((1+k^2)z'^2 + k^2 z^2\right)^{1/2} d\phi. \tag{10}$$

Without loss of generality we may further assume $\phi_1 = 0$.

### 1.3 Euler-Lagrangian equation.

A curve $z(\phi)$ which minimises $L$ (or, more generally, renders it stationary with respect to small variations) satisfies the Euler-Lagrange equation

$$\frac{d}{d\phi}\left(\frac{\partial f}{\partial z'}\right) = \frac{\partial f}{\partial z}, \tag{11}$$

where $f = ((1+k^2)z'^2 + k^2 z^2)^{1/2}$. It is convenient to divide $f$ by the constant $(1+k^2)^{1/2}$. Note that this leaves the Euler-Lagrange equation unchanged, and amounts to rescaling the length $L$ by a fixed factor. Let

$$a^2 = \frac{k^2}{1+k^2} = \frac{\tan^2\alpha}{1+\tan^2\alpha} = \frac{\tan^2\alpha}{\sec^2\alpha} = \sin^2\alpha,$$

so that

$$a = \sin\alpha. \tag{12}$$

Then we may take

$$f = (z'^2 + a^2 z^2)^{1/2}. \tag{13}$$

Since $f$ does not depend explicitly on $\phi$, we may use the alternative form of the Euler-Lagrange equation,

$$f - \frac{\partial f}{\partial z'} z' = c, \ \text{const.} \tag{14}$$

After some calculation, this gives

$$\frac{a^2 z^2}{(z'^2 + a^2 z^2)^{1/2}} = c. \tag{15}$$

To solve the Euler-Lagrange equation, first observe that a specific case is $z_1 = z_2$, i.e. the endpoints of the curve are on the same hight. Then (15) admits a trivial solution $z(\phi) = const$. However, it is in fact not a solution of the Euler-Lagrange equation – this can be verified directly – unless $c = 0$, hence $z = 0$! Indeed, consider the alternative form of the Lagrange equation (14) and differentiate it with respect to time. The result will be: either $z' = 0$, or $\frac{d}{dt} f_{z'} = f_z$, the Euler-Lagrange equation. Thus, the alternative form of the Lagrange equations can generate a "rubbish" solution $z = const$, which should always be checked with the Euler-Lagrange equation $\frac{d}{dt} f_{z'} = f_z$. It is easy to see that if $f$ is given by (13), the only constant solution of the Euler-Lagrange equation is $z = 0$, corresponding to the apex of the cone.

But there are other, less trivial solutions of (15), and they are minimisers. This will not be shown explicitly, but can be done using the unwrapping argument – after the cone having been unwrapped, the solutions we are about to find will become straight line segments that shall be glued along the edges of the sector, in the way described in the preamble.

Let us express $z'$ in terms of $z$ from (15):

$$\frac{dz}{d\phi} = \pm az\sqrt{\frac{a^2}{c^2}z^2 - 1}. \tag{16}$$

In the boundary conditions (5), without loss of generality, we can assume $z_1 \geq z_2$. Then "+" corresponds to curves which go up ($z$ increasing) in the anticlockwise (increasing-$\phi$) direction, and a "−" to curves which go up in the clockwise direction. It suffices to consider the "+" case, and we'll make the necessary remarks about the "−" case along the way after we see what's going on. In (16) variables separate, and it can be integrated as follows (using indefinite integrals for convenience)

$$\int \frac{dx}{z\sqrt{a^2z^2/c^2 - 1}} = a\phi + u_1, \tag{17}$$

where $u_1$ is some constant of integration. Recall that $\int \frac{dz}{z(z^2-1)^{1/2}} = \sec^{-1} z$, the inverse secant. Indeed, a trigonometric substitution $z = \sec u$ yields $dz = \frac{\sin u du}{\cos^2 u}$ and reduces the above integral to simply $\int du$. Similarly, $\int \frac{dx}{z\sqrt{a^2z^2/c^2 - 1}} = \sec^{-1}[(a/c)z]$.

Thus

$$z(\phi) = \frac{c}{a\cos(a\phi + u_1)}. \tag{18}$$

$c$ and $u_1$ are constants to be determined from the boundary conditions. The constant $u_1$ is clearly defined up to a multiple of $2\pi$, and in fact one can assume $|u| < \frac{\pi}{2}$. Adding $\pi$ to $u_1$ would negate the cosine and correspond to the choice of the "−" sign in (16). Also note that $u_1$ may be not allowed to change continuously over an interval of values of length $\pi$ and longer, because otherwise, given $\phi$, the denominator of (18) will zero at some point: e.g. the secant is only defined for $|a\phi + u_1| < \pi/2$.

### 1.4 Boundary conditions.

The quantities $c$ and $u_1$ are easily determined by the boundary conditions (5). It turns out that there may be more that one set of allowed values.

From (18), the boundary condition $z(\phi_1 = 0) = z_1$ implies that

$$c = az_1 \cos u_1, \tag{19}$$

so that (18) can be written as

$$z(\phi) = \frac{z_1 \cos u_1}{\cos(a\phi + u_1)}. \tag{20}$$

The boundary condition $z(\phi_2) = z_2$ implies that

$$z_2 = \frac{z_1 \cos u_1}{\cos(a\phi_2 + u_1)}. \tag{21}$$

With the identity $\cos(a\phi_2 + u_1) = \cos a\phi_2 \cos u_1 - \sin a\phi_2 \sin u_1$, this can be rearranged to give the following equation for $u_1$:

$$\tan u_1 = \frac{z_2 \cos(a\phi_2) - z_1}{z_2 \sin(a\phi_2)}. \tag{22}$$

Therefore,

$$u_1 = \tan^{-1}\left(\frac{z_2 \cos(a\phi_2) - z_1}{z_2 \sin(a\phi_2)}\right), \tag{23}$$

this ensures $|u_1| < \pi/2$. It also enables us to find $\cos u_1$ for (19), as a positive root of the quadratic equation $\tan^2 u_1 = \frac{1-\cos^2 u_1}{\cos^2 u_1}$, its left-hand side known from (22).

Recall that by (18) and (19),

$$z(\phi) = \frac{z_1 \cos u_1}{\cos(a\phi)\cos u_1 - \sin(a\phi)\sin u_1} = \frac{z_1}{\cos(a\phi) - \sin(a\phi)\tan u_1}, \tag{24}$$

therefore we finally get

$$z(\phi) = \frac{z_1 \cos u_1}{\cos(a\phi)\cos u_1 - \sin(a\phi)\sin u_1} = \frac{z_1}{\cos(a\phi) - \sin(a\phi)\frac{z_2\cos(a\phi_2)-z_1}{z_2\sin(a\phi_2)}}. \tag{25}$$

Let us analyse this solution. Recall that $a = \sin\alpha \in (0,1)$, hence $z(\phi)$ is $2\pi/a$-periodic in $\phi$, rather than $2\pi$-periodic. The angle $\phi$ varies continuously from 0 to $\phi_2$, and instead of considering $\phi_2$ modulo $2\pi$, let us allow it to have real values. What would it mean to say that we have a solution $z(\phi)$ given by (21) or (25) with $\phi_2 \in [2\pi, 4\pi]$? Such a curve would start out at the first terminal point $A$, and before coming to the second terminal point $B$ it would wrap around the cone! Hence, if (25) would make sense for the solutions $z(\phi)$ with $\phi_2 \in [2\pi m, 2\pi(m+1)]$, these solutions will wrap counterclockwise around the cone $m$ times (and will do it clock-wise for negative $m$, thus allowing us to take into account the "$-$" sign case mentioned above.)

The winding number $m$ is bounded however. Indeed, in order that the solution $z(\phi)$ make sense, we have to ensure that for all $\phi \in (0, \phi_2)$ the denominator in (25) or (18) is not zero, i.e.

$$|a\phi + u_1| < \pi/2, \quad \forall \phi \in [0, \phi_2]. \tag{26}$$

In other words, we must have

$$-\pi/2 - u_1 < a\phi < \pi/2 - u_1. \tag{27}$$

Hence, if the terminal point $B$ is positioned at the azimuth $\phi_2^0 \in (-\pi, \pi]$, the number of distinct solutions $N$ cannot exceed the number of points $\phi_2^0 + 2\pi m$, for integer $m$, that fall into the interval $\left(\frac{-\pi/2-u_1}{2a}, \frac{\pi/2-u_1}{2a}\right)$, of length $\pi/2$. The number of such points, by inspection, equals

$$N_+ = 1 + \left[\frac{1}{2a}\right]$$

where $[x]$ denotes the integer part of $x$ (i.e., the largest integer not smaller than $x$). Indeed, there is always at least one geodesic. On the other hand, from (27) one can see that no matter what $u_1$ is (it depends on the boundary conditions, hence the winding number, the worst case being $u_1 = 0$), the angle $\phi$ can change continuously from zero in either positive or negative direction up to any value that is smaller than $\frac{\pi}{2a}$, so that the inequality (27) be satisfied. Hence, no matter what the initial conditions, the number of distinct secant solutions – i.e. those in the form (21) – is bounded from below by

$$N_- = 1 + \left[\frac{1}{4a}\right].$$

Finally then, for the number $N$ of secant solutions we have the estimate

$$1 + \left[\frac{1}{4a}\right] \leq N \leq 1 + \left[\frac{1}{2a}\right]. \tag{28}$$

**Example.** For simplicity, let us take the endpoints of the curve to coincide, so that the curve we are looking for connects a chosen point to itself, i.e., is closed. Let us identify the secant solutions given by (21) or (25) and having different winding numbers.

We take

$$(z_1, \phi_1) = (1, 0), \quad (z_2, \phi_2) = (1, 2\pi m), \tag{29}$$

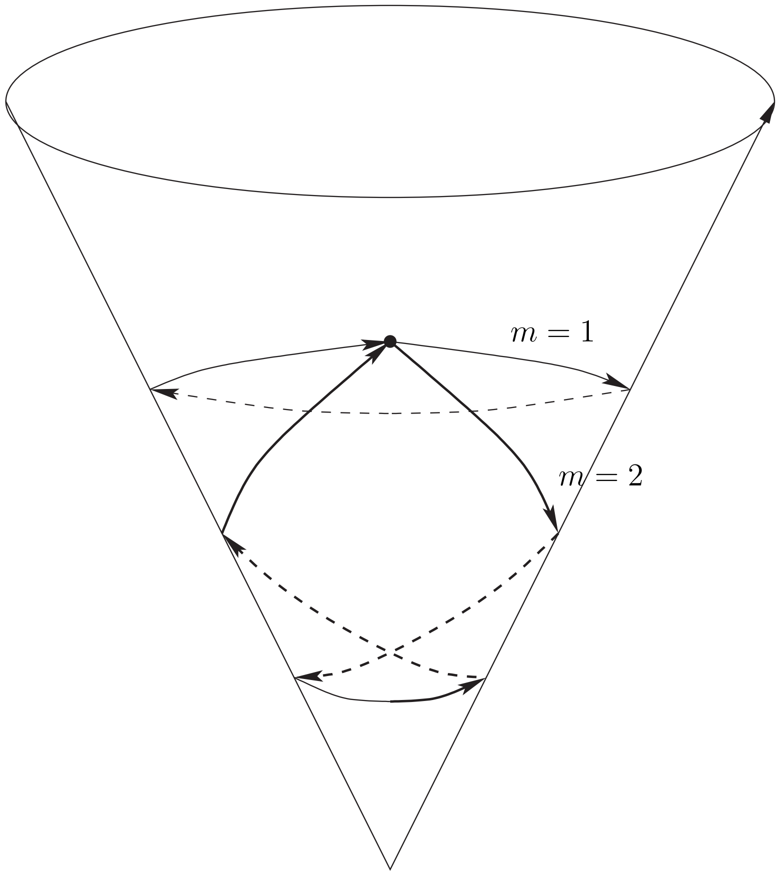
where $m$ is an integer. Then $\phi_2 = 2\pi m$, and (23) simplifies to

$$\begin{aligned} u_1 &= \tan^{-1}\left(\frac{\cos 2\pi ma - 1}{\sin 2\pi ma}\right) = \tan^{-1}\left(\frac{-2\sin^2 \pi ma}{2\sin \pi ma \cos \pi ma}\right) \\ &= \tan^{-1}(-\tan \pi ma)) \;=\; -\pi ma \end{aligned} \tag{30}$$

as long as $|\pi ma| < \pi/2$, which will indeed be the case. Each value of $m$ leads to a different secant solution $z^{(m)}(\phi)$:

$$z^{(m)}(\phi) = \frac{\cos \pi ma}{\cos(a\phi - \pi ma)}, \quad 0 \leq \phi \leq 2\pi m. \tag{31}$$

which satisfies the Euler-Lagrange equation, and $m$ determines the number of times the curve winds around the cone ($m$ is positive for anticlockwise windings, and negative for clockwise). The case $m = 0$ corresponds to the trivial curve of zero length consisting of the single point $z = 1$, $\phi = 0$. For $m \neq 0$, the curves given by (31) clearly have length greater than zero, and provide minimisers for length in the class of closed curves that wind $m$ times around the cone. (The fact that they are minimisers, but not just stationary points of the functional $L$ either requires additional analytic argument with is beyond the scope of this exposition, or follows from the "unwrapping" technique and is left to the reader.)




As $\phi$ varies between 0 and $2\pi m$, $a(\phi - m\pi)$ varies between $-m\pi a$ and $m\pi a$. If this range were to include $\pm\pi/2$, then $\cos(a(\phi - m\pi))$ would vanish, and $z^{(m)}(\phi)$ would diverge. Therefore, for $z^{(m)}(\phi)$ to be well defined, we require that $|m\pi a| < \pi/2$, (which justifies the last step in (30). I.e.,

$$|m| < \frac{1}{2a} = \frac{1}{2\sin\alpha}. \tag{32}$$

Observe, however, that curves corresponding to opposite values of $m$ actually coincide – this is the artefact of $\phi_1 = \phi_2$ modulo $2\pi$.

Thus, the number of distinct geodesics connecting the point $A:\ (z,\phi)=(0,1)$ to itself is

$$N = 1 + \left[\frac{1}{2\sin\alpha}\right],$$

which indeed equals the upper bound in the general estimate (28).

It is interesting to consider what happens to a given solution $z^{(m)}(\phi)$ as $\alpha$ changes, so the cone is widened or narrowed. By (28), the narrower the cone, the more secant solutions with different winding numbers there are in general. Let us now fix the winding number $m$, and suppose $\alpha$ is such that (32) is satisfied, so we have a secant solution $z^{(m)}(\phi)$, given by (31). Note that the highest point (maximum value of $z$) along $z^{(m)}(\phi)$ is at $\phi=0$ and $\phi=2\pi m$, ie the endpoint, where $z=1$, while the lowest point is at $\phi=m\pi$, where $z=\cos m\pi a<1$. As $\alpha$ increases, so does $a=\sin\alpha$, and the lowest point sinks lower. When $a=1/(m\pi)$, the curve touches the apex of the cone, and for larger values of $\alpha$ is no longer defined. So in fact, we have now succeeded in making the inequality (32) non-strict, which would correspond to $u_1=-\pi/2$ and described the limit case of geodesics passing through the apex, despite the angle $\phi$ is not defined there!