# GEODESICS AND THEIR MINIMIZATION PROPERTIES

KIRIL DATCHEV

## 1. Shortest paths in Euclidean space

Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be a continuous function. The *length* of $\gamma$ is

$$L = \sup \sum_{j=1}^{n} |\gamma(t_j) - \gamma(t_{j-1})|,$$

where the sup is taken over all partitions $a = t_0 < t_1 < \cdots < t_n = b$ of the interval $[a,b]$. If $\gamma$ is continuously differentiable, then

$$L = \int_a^b |\gamma'(t)|dt; \tag{1}$$

see [Ru, Theorem 6.27]. If $p$ and $q$ are two points in $\mathbb{R}^n$, then the line segment joining them has length strictly less than that of any other curve from $p$ to $q$. This follows from the *triangle inequality*: if $p$, $q$, and $r$ are distinct points in $\mathbb{R}^n$ such that $r$ does not lie on the segment from $p$ to $q$, then

$$|p-q| < |p-r| + |r-q|. \tag{2}$$

The proof of this simple statement is perhaps surprisingly complicated. It is Proposition 20 in Euclid [Eu], and relies on most of what comes before it there (although not the parallel postulate). We proceed algebraically, starting with the easier result that if $p$, $q$ and $r$ are any points in $\mathbb{R}^n$, then

$$|p-q| \le |p-r| + |q-r|. \tag{3}$$

To prove (3), put $x = p - r$ and $y = r - q$ and rewrite it as

$$|x+y| \le |x| + |y|, \tag{4}$$

which is proved by a reduction to *Cauchy's inequality*:

$$|x \cdot y| \le |x||y|; \tag{5}$$

see [Sp, Theorem 1-1] or [Ta, Proposition 1.2.2],

1.1. **Exercise.** Prove (2) by showing that if $|p-q| = |p-r| + |r-q|$, then $r = \alpha p + (1-\alpha) q$ for some $\alpha \in (0,1)$. You may want to follow these steps:

a) Rewrite $|p-q| = |p-r| + |q-r|$ using $x$ and $y$ in the manner of (4)
b) Reduce to an analogue of (5).
c) Use the strategy of the proof of [Ta, Proposition 1.2.2] to find an explicit linear dependence equation between $x$ and $y$ implied by the analogue of (5) found in b) above.
d) Use the equation found in c) above to combine the equations defining $x$ and $y$ into an equation for $r$ of the desired form.

*Date*: March 17, 2021. Please email any comments or corrections to kdatchev@purdue.edu.

## 2. Shortest paths in curved space

2.1. **Manifolds.** We say $M \subset \mathbb{R}^m$ is a $C^\infty$ *n-dimensional manifold* (or *surface*) if, for every $p \in M$ there are open sets $W \subset \mathbb{R}^m$ and $U \subset \mathbb{R}^n$ and a $C^\infty$ function $\varphi\colon U \to W$ such that $p \in W$, $U = \varphi^{-1}(M \cap W)$, $\varphi\colon U \to M \cap W$ is invertible with continuous inverse, and $D\varphi$ has full rank (i.e. is one-to-one or injective) at all points of $U$. We call $M \cap W$ a *coordinate patch*, and $\varphi$ a *coordinate chart*.

2.2. **Examples.**

(1) Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a $C^\infty$ function. Then the graph of $M$ is a $C^\infty$ $n$-dimensional manifold in $\mathbb{R}^{n+1}$. The whole manifold is a coordinate patch with coordinate chart $\varphi(x) = (x, f(x))$.

(2) Let $a > 0$, and let $M = \varphi(\mathbb{R})$, where $\varphi\colon \mathbb{R} \to \mathbb{R}^2$ be given by

$$\varphi(\alpha) = (a\cos\alpha, a\sin\alpha).$$

Then $M$ is a circle. If we take $U = I$, where $I$ is an interval of length $2\pi$, then $\varphi(U)$ is a coordinate patch and the restriction of $\varphi$ to $U$ is a coordinate chart. If the length of $I$ is greater than $2\pi$, then the restriction of $\varphi$ is not one-to-one, and if the length equals $2\pi$ then the restriction is invertible but the inverse is not continuous.

(3) The above example works the same way if instead $\varphi\colon \mathbb{R} \to \mathbb{R}^3$ is given by

$$\varphi(\alpha) = (a\cos\alpha, a\sin\alpha, 0).$$

(4) Let $a > b > 0$, and let $M = \varphi(\mathbb{R}^2)$, where $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^3$ is given by

$$\varphi(\alpha,\beta) = (a\cos\alpha, a\sin\alpha, 0) + (b\cos\alpha\cos\beta, b\sin\alpha\cos\beta, b\sin\beta). \tag{6}$$

Then $M$ is a *torus* or *inner tube* as in Figure 3.2.3 of [Ta] or Figure 5-4 of [Sp]. If we take $U = I \times J$, where $I$ and $J$ are both open intervals of length less than $2\pi$, then $\varphi(U)$ is a coordinate patch and the restriction of $\varphi$ to $U$ is a coordinate chart.

2.3. **Exercise.** Check that if $\varphi$ given by (6) then $D\varphi(\alpha,\beta)$ has full rank for any real $\alpha$ and $\beta$.

2.4. **Curves on manifolds.** Let $\gamma\colon [a,b] \to M$ be a $C^\infty$ curve. Then its length is given by (1), $L = \int_a^b |\dot\gamma|$, where we use $\dot\gamma$ to denote the derivative with respect to $t$. To express this in terms of coordinates, we consider $[a,b]$ short enough that $\gamma([a,b])$ is contained in a coordinate patch, and let $x(t) = \varphi^{-1}(\gamma(t))$. Then $\dot\gamma(t) = D\varphi(x(t))\dot x(t)$, and

$$|\dot\gamma(t)|^2 = |D\varphi(x(t))\dot x(t)|^2.$$

We write out the right hand side in terms of components as

$$\begin{aligned}\Big(D\varphi(x(t))\dot x(t)\Big)\cdot\Big(D\varphi(x(t))\dot x(t)\Big) &= \sum_{\ell=1}^m \Big(\sum_{j=1}^n \frac{\partial\varphi_\ell(x(t))}{\partial x^j}\dot x^j(t)\Big)\Big(\sum_{k=1}^n \frac{\partial\varphi_\ell(x(t))}{\partial x^k}\dot x^k(t)\Big)\\ &= \sum_{j,k=1}^n g_{jk}(x(t))\dot x^j(t)\dot x^k(t)\end{aligned}$$

where

$$g_{jk}(x(t)) = \sum_{\ell=1}^m \frac{\partial\varphi_\ell(x(t))}{\partial x^j}\frac{\partial\varphi_\ell(x(t))}{\partial x^k};$$

in other words, $g_{jk}$ is the dot product of the $j$th and $k$th columns of $D\varphi$. Note that here, and below, we are using the superscript notation $x^j$ to denote the $j$th component of $x$, not a power of

$x$. We abbreviate this further using the *summation convention* that repeated indices are summed over, as

$$|D\varphi(x(t))\dot{x}(t)|^2 = g_{jk}(x(t))\dot{x}^j(t)\dot{x}^k(t) = g_{jk}\dot{x}^j\dot{x}^k.$$

Thus the $g_{jk}$ are the entries of the matrix $G := (D\varphi)^T D\varphi$, i.e. the product of the derivative matrix $D\varphi$ with its transpose, and we call this matrix the *metric tensor* on $M$ in the coordinate chart $\varphi(U)$. To define geodesics and prove they locally minimize distances we will only need the following properties of the metric tensor: $G(x)$ is symmetric and positive definite for all $x$, and its entries $g_{jk}(x)$ are $C^\infty$ functions of $x$.

2.5. **Example.** In Example 2.2 (1), the matrix of $D\varphi$ is $\binom{I}{Df}$, where $I$ is the $n \times n$ identity matrix and $Df$ is the $1 \times n$ gradient matrix $(\partial_1 f, \dots, \partial_n f)$. Thus $G = (D\varphi)^T D\varphi = I + (Df)^T Df$, or

$$g_{jk} = \delta_{jk} + \partial_j f \partial_k f,$$

where $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ if $j \neq k$. In other words the metric tensor in these coordinates is the sum of the identity matrix and a scaled orthogonal projection onto the gradient of $f$.

2.6. **Exercise.** Find $g_{jk}$ for the helicoid[1] parametrized by

$$\varphi\colon (0,1) \times \mathbb{R} \to \mathbb{R}^3, \qquad \varphi(x^1, x^2) = (x^1 \cos x^2, x^1 \sin x^2, x^2),$$

and for the sphere parametrized by

$$\varphi\colon \mathbb{R} \times (0,\pi) \to \mathbb{R}^3, \qquad \varphi(x^1, x^2) = (\cos x^1 \sin x^2, \sin x^1 \sin x^2, \cos x^2).$$

2.7. **Definition of a geodesic.** We now define a geodesic in terms of such coordinates. For $1 \le j,\, k \le n$, let $g_{jk} \in C^\infty(\mathbb{R}^n)$ be such that for all $x$ we have $g_{jk}(x) = g_{kj}(x)$ and $g_{jk}(x)v^j v^k > 0$ for any nonzero $v = (v^1, \dots, v^n) \in \mathbb{R}^n$, where we use the summation convention that repeated indices are summed over, i.e. $g_{jk}v^j v^k = \sum_{j=1}^n \sum_{k=1}^n g_{jk}v^j v^k$. If $x = (x^1, \dots, x^n)\colon [a,b] \to \mathbb{R}^n$ is a $C^\infty$ curve, its *speed* with respect to $g$ at time $t$ is given by $\sqrt{g_{jk}(x(t))\dot{x}^j(t)\dot{x}^k(t)}$, and its *length* with respect to $g$ is given by

$$L = L_g(x) = \int_a^b \sqrt{g_{jk}\dot{x}^j\dot{x}^k}dt. = \int_a^b \sqrt{g_{jk}(x(t))\dot{x}^j(t)\dot{x}^k(t)}dt.$$

We wish to find the curves which minimize this length. We begin by looking at the simpler *energy functional*, which is defined by

$$E = E_g(x) = \frac{1}{2}\int_a^b g_{jk}\dot{x}^j\dot{x}^k.$$

To look for a minimizer, we take $x_s(t)$ such that $x_s(a)$ and $x_s(b)$ are both independent of $s$, and differentiate with respect to $s$:

$$\partial_s E = \int_a^b \left(g_{jk}\dot{x}^j\partial_s\dot{x}^k + \tfrac{1}{2}\partial_\ell g_{jk}\dot{x}^j\dot{x}^k\partial_s x^\ell\right) dt.$$

Then we integrate by parts in the first term (removing a $t$ derivative from $\partial_s\dot{x}^k$) and swap the $k$ and $\ell$ indices in the last term to get

$$\partial_s E = \int_a^b \left(-g_{jk}\ddot{x}^j - \partial_\ell g_{jk}\dot{x}^\ell\dot{x}^j + \tfrac{1}{2}\partial_k g_{j\ell}\dot{x}^j\dot{x}^\ell\right)\partial_s x^k dt.$$

[1] See https://en.wikipedia.org/wiki/Helicoid

If $x_0$ is a critical point of the energy functional, then $\partial_s E|_{s=0} = 0$. If this is the case for any variation $x_s$, then we obtain the system of geodesic equations

$$-g_{jk}\ddot{x}^j - \partial_\ell g_{jk}\dot{x}^\ell\dot{x}^j + \tfrac{1}{2}\partial_k g_{j\ell}\dot{x}^j\dot{x}^\ell = 0, \qquad \text{for } k = 1,\dots,n. \tag{7}$$

Any solution to (7) is called a *geodesic* with respect to $G$. Solutions are guaranteed to exist by the existence and uniqueness theorem for ordinary differential equations; more precisely, given any initial conditions $p$ and $v$ in $\mathbb{R}^n$, for some $T > 0$ there is a unique $C^\infty$ function $x\colon (-T,T) \to \mathbb{R}^n$ such that (7) holds and $x(0) = p$ and $\dot{x}(0) = v$.

We will see shortly that geodesics locally minimize length. For now, observe that differentiating $g_{jk}\dot{x}^j\dot{x}^k$ with respect to $t$ and plugging in (7) shows that all geodesics have constant speed. Note also that if $x(t)$ is a geodesic, so is $x(\lambda t)$ for any real $\lambda$.

2.8. **Examples.** The simplest example is free space, where $g_{jk} \equiv \delta_{jk}$, and the geodesic equations become $\ddot{x} = 0$. Thus the geodesics are lines traversed at constant speed; $x(t) = p + vt$.

In the torus example above, replacing $\alpha$ by $x^1$ and $\beta$ by $x^2$ in (6), we have

$$D\varphi = \begin{pmatrix} -\sin x^1(a + b\cos x^2) & -b\cos x^1 \sin x^2 \\ \cos x^1(a + b\cos x^2) & -b\sin x^1\sin x^2 \\ 0 & b\cos x^2 \end{pmatrix}, \quad G = (D\varphi)^T D\varphi = \begin{pmatrix} (a+b\cos x^2)^2 & 0 \\ 0 & b^2 \end{pmatrix}.$$

Thus $g_{11} = (a + b\cos x^2)^2$, $g_{12} = g_{21} = 0$, and $g_{22} = b^2$. The geodesic equations are

$$-(a + b\cos x^2)^2\ddot{x}^1 - 2(a + b\cos x^2)(-b\sin x^2)\dot{x}^1\dot{x}^2 = 0,$$

$$-b^2\ddot{x}^2 + (a + b\cos x^2)(-b\sin x^2)\dot{x}^1\dot{x}^1 = 0.$$

The general solutions of this system do not have a simple form, but those with $\ddot{x} \equiv 0$ do, namely $x(t) = p + tv$, where either $v^1 = 0$, or $(p^2, v^2) = (m\pi, 0)$.

2.9. **Exercises.**

(1) Find the geodesic equations for the ellipsoid parametrized by

$$\varphi\colon \mathbb{R}\times(0,\pi) \to \mathbb{R}^3, \qquad \varphi(x^1,x^2) = (a\cos x^1\sin x^2, a\sin x^1\sin x^2, b\cos x^2),$$

where $a > 0$ and $b > 0$ are given constants. Find all $(p, v)$ for which $x(t) = p + vt$ is a geodesic.

(2) Let $R > 0$ be given, and let $p$ and $q$ be two points on a circular cylinder of radius $R$. Use the parametrization $\varphi(\theta, z) = (R\cos\theta, R\sin\theta, z)$, where $\varphi(0,0) = p$ and $\varphi(\theta_0, z_0) = q$ for some $(\theta_0, z_0) \in [-\pi,\pi]\times\mathbb{R}$, to find the lengths of all possible geodesics from $p$ to $q$.

2.10. **Normal coordinates.** In the above Exercise one finds that curves of the form $x(t) = p+vt$ solve the geodesic equations when $p^2 = \pi/2$ and $v^2 = 0$. When $a = b$, the ellipsoid is a sphere, and it is possible to choose the coordinates on the sphere in such a way that any preassigned $\gamma(0)$ and $\gamma'(0)$ are given in coordinates by such a $(p, v)$. This proves that all geodesics on a sphere are great circles traversed at constant speed, and it is also an example of the simplifying power of well-chosen coordinates.

To prove that geodesics locally minimize length on a more general manifold, we introduce *normal coordinates* near 0; these are coordinates which give the metric the simplest possible form near 0. We define them in two steps.

First let $A$ be a linear map $\mathbb{R}^n \to \mathbb{R}^n$ which takes the standard basis vectors $e_1,\dots,e_n$ to a basis $u_1,\dots,u_n$ which is orthonormal with respect to the metric at 0, i.e. such that

$$\Big[G(0)u_k\Big]\cdot u_\ell = g_{ij}(0)u_k^i u_\ell^j = \delta_{k\ell}.$$

This gives rise to new coordinates $\tilde{x}$ with a corresponding metric tensor $\tilde{G}$, such that

$$\Big[\tilde{G}(\tilde{x})\dot{\tilde{x}}\Big] \cdot \dot{\tilde{x}} = \Big[G(x)\dot{x}\Big] \cdot \dot{x},$$

where we put $\tilde{x} = A^{-1}x$ and $\tilde{G}(\tilde{x}) = A^T G(x) A$. Then $\Big[\tilde{G}(0)e_k\Big] \cdot e_\ell = \Big[G(0)u_k\Big] \cdot u_\ell = \delta_{k\ell}$, so that $\tilde{G}(0) = I$, and hence in the new coordinates we have

$$\tilde{g}_{ij}(0) = \delta_{ij}. \tag{8}$$

Second, let $\mathrm{Exp}(v) = \gamma_v(1)$, where $\gamma_v$ is the geodesic such that $\gamma_v(0) = 0$ and $\dot{\gamma}_v(0) = v$. This may not be defined for all $v$ (because the geodesic equation may not be solvable up to $t = 1$ for all $v$), but the existence, uniqueness, and regularity theorem for ordinary differential equations guarantees that $\mathrm{Exp}(v)$ is defined for $v$ small enough and that $\mathrm{Exp}\colon B(c) \to \mathbb{R}^n$ is a $C^\infty$ map for some $c > 0$, where $B(c)$ is the open ball centered at 0 with radius $c$. The derivative of Exp at 0 is the identity, because

$$\mathrm{Exp}(\lambda v) = \gamma_{\lambda v}(1) = \gamma_v(\lambda) = \gamma_v(0) + \lambda\dot{\gamma}_v(0) + r(\lambda, v) = \lambda v + r(\lambda, v),$$

where $|r(\lambda, v)|/\lambda \to 0$ as $\lambda \to 0$. Hence, by the inverse function theorem, after possibly shrinking $c$, the map $\mathrm{Exp}\colon B(c) \to \mathrm{Exp}(B(c))$ is invertible with $C^\infty$ inverse.

In these coordinates we write the metric as $H$, its entries as $h_{jk}$, and points as $y = (y^1, \dots y^n)$ so that curves are $y(t) = (y^1(t), \dots, y^n(t))$. If $y(t)$ is a geodesic and $y(0) = 0$, $\dot{y}(0) = v$, then $y(t) = tv$ since $\mathrm{Exp}(tv)) = \gamma_{tv}(1) = \gamma_v(t)$. Further, since geodesics have constant speed, we have

$$h_{jk}(v)v^j v^k = h_{jk}(y(1))\dot{y}^j(1)\dot{y}^k(1) = h_{jk}(y(0))\dot{y}^j(0)\dot{y}^k(0) = h_{jk}(0)v^j v^k.$$

We simplify the right hand side using

$$H = (D\,\mathrm{Exp})^T \circ \tilde{G} \circ D\,\mathrm{Exp};$$

so by (8) and using $D\,\mathrm{Exp}(0) = I$ we have $h_{jk}(0) = \delta_{jk}$, and hence

$$h_{jk}(v)v^j v^k = \delta_{jk} v^j v^k = |v|^2, \tag{9}$$

for any $v \in \mathbb{R}^n$ such that $|v| < c$. Thus we know $h$ on radial vectors. The Gauss Lemma says that radial vectors are $h$-perpendicular to those vectors to which they are Euclidean-perpendicular; i.e.

$$v \cdot w = 0 \quad \Longrightarrow \quad \big[H(v)v\big] \cdot w = h_{jk}(v)v^j w^k = 0. \tag{10}$$

To prove (10), let

$$y_s(t) = \frac{v + sw}{|v + sw|}t, \qquad E(y_s) = \frac{1}{2}\int_0^b h_{jk}(y_s(t))\dot{y}_s^j(t)\dot{y}_s^k(t)dt,$$

where $b = |v|$. Then, as before, we get

$$\partial_s E = \int_0^b \Big(h_{jk}\dot{y}^j \partial_s \dot{y}^k + \tfrac{1}{2}\partial_\ell h_{jk}\dot{y}^j\dot{y}^k\partial_s y^\ell\Big)\,dt,$$

and integrating by parts, and swapping the $k$ and $\ell$ indices in the last term gives

$$\partial_s E = h_{jk}(y(b))\dot{y}^j(b)\partial_s y^k(b) + \int_0^b \Big(-h_{jk}\ddot{y}^j - \partial_\ell h_{jk}\dot{y}^\ell\dot{y}^j + \tfrac{1}{2}\partial_k h_{j\ell}\dot{y}^j\dot{y}^\ell\Big)\,\partial_s y^k dt.$$

Combining this with $\partial_s E = 0$ (because $y_s$ has unit speed for every $s$) and observing that the integrand vanishes because $y$ solves the geodesic equation, and plugging in $s = 0$, gives (10).

Let $y(t)$ be any path in $B(c) \setminus 0$, and let $r(t) = |y(t)|$. Then $y(t) = r(t)\omega(t)$, where $\omega(t) = y(t)/|y(t)|$, and

$$\Big[H(y(t))\dot{y}(t)\Big] \cdot \dot{y}(t) = \Big[H\dot{y}\Big] \cdot \dot{y} = \dot{r}^2\Big[H\omega\Big] \cdot \omega + r^2\Big[H\dot{\omega}\Big] \cdot \dot{\omega} = \dot{r}^2 + r^2\Big[H\dot{\omega}\Big] \cdot \dot{\omega} \geq \dot{r}^2.$$

where for the second equality we used the fact that (10) implies $[H\omega] \cdot \dot{\omega} = 0$, for the third equality we used (9), and the inequality is strict if and only if $\dot{\omega} \neq 0$ because $H$ is positive definite.

Now let $p \in B(c)$, let $y\colon [a,b] \to B(c)$ be any smooth path with $y(a) = 0$ and $y(b) = p$. The length is

$$\int_a^b \sqrt{\Big[H(y(t))\dot{y}(t)\Big] \cdot \dot{y}(t)}dt \geq \int_a^b |\dot{r}| \geq \int_a^b \dot{r} = r(b) - r(a) = |p|,$$

with equality if and only if $\omega$ is constant and $r$ is monotonic, i.e. if and only if $y(t)$ monotonically parametrizes the line segment from 0 to $p$. Thus a parametric curve $\gamma$ from 0 to $p$ always has length $\geq$ the length of the geodesic curve from 0 to $p$, and the lengths are equal if and only $\gamma$ monotonically parametrizes the geodesic curve.

2.11. **Exercises.**

(1) Let $c > 0$ be given, and let $G(x) = f(|x|^2)I$, where $f\colon [0,c) \to (0,\infty)$ is $C^\infty$, and let $p \in B(0,c)$ be given. Mimic the above proof to show that if $x\colon [0,b] \to B(0,c)$ is $C^\infty$ with $x(0) = 0$, $x(b) = p$, and $|x(t)| > 0$ for $t > 0$, then the length of $x$ with respect to $G$ is $\geq$ the length of the line segment from 0 to $p$ with respect to $G$, with equality if and only if $x$ monotonically parametrizes the line segment.

(2) Define *stereographic coordinates* on the unit sphere $\mathbb{S}^2$ by the map $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^3$ which takes a point $U$ on the plane through the equator to the point on the unit sphere which is collinear with $U$ and with the north pole $N = (0,0,1)$; see Figure 1. Show that

$$\varphi(x) = \Big(\frac{2x}{|x|^2+1}, \frac{|x|^2-1}{|x|^2+1}\Big),$$

and $\varphi$ is a coordinate chart with coordinate patch $\varphi(\mathbb{R}^2) = \mathbb{S}^2 \setminus N$. Show that

$$G(x) = \frac{1}{(1+\frac{1}{4}|x|^2)^2}I,$$

and hence that the result of part (1) is applicable here. Can you get a similar result for an ellipsoid as in part (1) of Exercise 2.9? More generally, for an $n$-dimensional sphere of radius $R$, the same construction gives

$$G(x) = \frac{1}{(1+\frac{\alpha}{4}|x|^2)^2}I,$$

where $\alpha = 1/R^2$ is the *curvature* of the sphere. This last formula is the sole displayed equation of Riemann's very important habilitation lecture on geometry [Ri]; it is a general formula for manifolds of constant curvature. In particular it works also when $\alpha \leq 0$, although when $\alpha < 0$ the domain of $\varphi$ is $B(2|\alpha|^{-1/2}) \subset \mathbb{R}^n$.

(3) Show that $w = \frac{z-i}{z+i}$ maps the upper half plane $\{z \in \mathbb{C} \mid \mathrm{Im} z > 0\}$ bijectively to the unit disk $\{w \in \mathbb{C} \mid |w| < 1\}$. With $z = y_1 + iy_2$ and $w = x_1 + ix_2$, show that the metric $G(x) = 4(1-|x|^2)^{-2}I$ on the unit disk corresponds in the $y$ coordinates to $H(y) = y_2^{-2}I$. These are the two best coordinate systems for the hyperbolic plane (corresponding to the case $\alpha = -1$ of part (2) above), called the upper half plane model and the unit disk model. See Figure 2 for an artist's rendition. Geodesic arcs are arcs of circles that intersect the boundary of the space orthogonally; see Proposition 2.3 of [Bo].

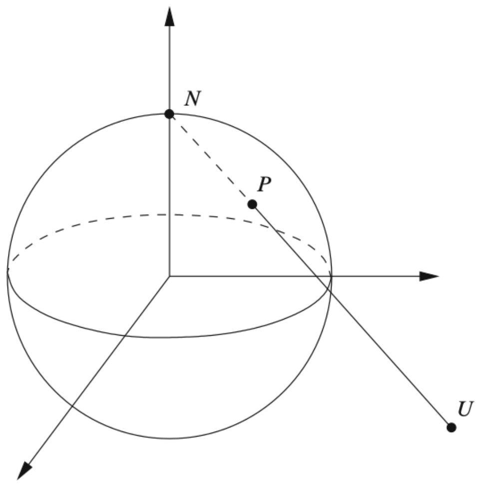


FIGURE 1. A depiciton of stereographic projection on the sphere, adapted from Figure 3.2 of [Le]. The map $\varphi$ takes a point $U$ in the horizontal plane to the point $P$ on the unit sphere which is collinear with $U$ and the north pole $N$.

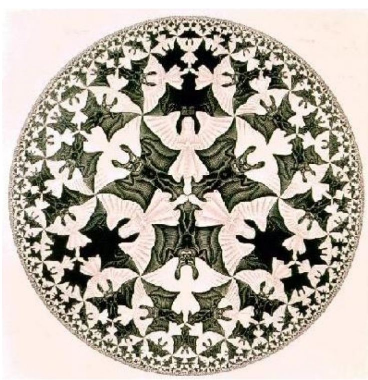

FIGURE 2. M. C. Escher's 1960 woodcut *Heaven and Hell*, and a mapping of it to the upper half plane model by Arkadiusz Jadczyk (`http://arkadiusz-jadczyk.eu/blog/2017/04/24/`). The angels and devils all have the same size with respect to the hyperbolic metric.

## 3. Further discussion and references

In Section 1 we showed that between any two points there is a unique geodesic, and that any other path between those two points has strictly greater length. In Section 2 we showed, following Sections 3.2 and 6.1 of [Ta], that for any point on any manifold, the same holds in a sufficiently small neighborhood of that point.

On the manifold as a whole the situation can be quite different. For example, on a sphere, if two points are antipodal, then there are infinitely many geodesics between them, all of the same length. Between non-antipodal points there are geodesics of different lengths. Nevertheless, minimizing curves are always geodesics (see [Mi, Corollary 10.7] or [Le, Theorem 6.6]), and on a complete manifold there is a minimizing curve between any two points (see [Mi, Corollary 10.7] or [Le, Corollary 6.15]).

The books of Milnor [Mi] and Lee [Le] also contain much more information about geodesics, and further references. Ratcliffe's book [Ra] gives a broader introduction to the topic, beginning with Euclid and discussing general metric spaces, before delving deeply into hyperbolic geometry. See in particular [Ra, Chapter 1] for more on geodesics in $\mathbb{R}^n$.

## References

[Bo] David Borthwick *Spectral Theory of Infinite-Area Hyperbolic Surfaces*, Second Edition, 2016.

[Eu] Euclid, *The Elements*, 300 BC, Translated by Sir Thomas L. Heath, 1925.

[Le] John M. Lee, *Riemannian Manifolds*, Springer Graduate Texts in Mathematics 176, 1997.

[Mi] J. Milnor, *Morse Theory*, Annals of Mathematics Studies 51, 1969.

[Ra] John G. Ratcliffe, *Foundations of Hyperbolic Manifolds*, Second Edition, Springer Graduate Texts in Mathematics 149, 2005.

[Ri] Bernhard Riemann, *On the hypotheses which lie at the basis of geometry*, 1854. http://www.cs.jhu.edu/~misha/ReadingSeminar/Papers/Riemann54.pdf

[Ru] Walter Rudin, *Principles of Mathematical Analysis*, Third Edition, 1976.

[Sp] Michael Spivak, *Calculus on Manifolds*, 1965.

[Ta] Michael E. Taylor, *Introduction to Analysis in Several Variables*, AMS Sally Series of Pure and Applied Undergraduate Texts 46, 2020. Preprint available online at https://mtaylor.web.unc.edu/wp-content/uploads/sites/16915/2018/04/analmv.pdf.