Navigating the Negative Curvature of Google Maps

Yuliy Baryshnikov[*]
Robert Ghrist[†]

February 1, 2023

Introduction

With more and more of one's life spent in front of a screen, working one application after another, the quality of the design of user interface becomes increasingly important.

Poor design is felt viscerally: navigating through clunky interfaces that force upon us unnecessary work registers as an annoyance unto impediment. Good design remains largely unnoticed: such feels natural, even if it is the end-product of hard work, keen insight, and relentless focus. Sometimes, good design is seredipitous, dictated by the very nature of the task. This note explores the mathematical underpinning of the happy accident behind one such elegant, powerful, and rarely noticed user interface in Google Maps.

(Historical note: Google Maps, launched as a web application in 2005, emerged as the result of acquisition by Google of Where2, an Australian startup. In 2007 Google Maps became the default map application on the newly-released iPhone, with the Android version launched a year later. In 2012 Apple developed its own version of the application, Apple Maps. As the technology matured, other apps emerged: they all share key interface features with Google Maps.)

Google Maps is installed on the smartphones of billions of people, who use it regularly. Besides its unquestionable utility, it is remarkably easy to use. One aspect of this user-friendliness – the ease of navigation in the *viewport space* – has a deep mathematical basis for why it is pleasant to operate.

A *viewport* is simply what one sees on the screen – a rectangular fragment of a map. When you pinch or expand or flick to scroll your screen, transforming the viewport, you are in fact navigating through a geometric *viewport space*. Unlike actual navigation of physical space, navigation in viewport space is swift and intuitive (not to mention ecologically harmless). It is perhaps the most convenient user interface one regularly interfaces with.

What makes this so convenient? The reason – the kernel of this note – is that viewport manipulation on a phone is in fact navigation along geodesics in a three-dimensional hyperbolic space.

Viewport space

Google Maps (unlike Google Earth, which manifestly accepts that the Earth's surface is round) uses a version of the Mercator projection. Avoiding details, it suffices to say that in the computational pipe, the surface of Earth is projected to a cylinder whose axis is aligned with the Earth's rotational axis. This cylinder is then unrolled into a planar image (which is thus periodic east-west). A viewport in

[*]https://ymb.web.illinois.edu
[†]https://www2.math.upenn.edu/ ghrist/

this setting is a rectangle aligned with the coordinate axes of the resulting plane, with a fixed aspect ratio.[1]

A natural way to think about a viewport is to imagine viewing the planar map below through the phone's camera while hovering above it at a certain height with a fixed zoom setting. With this in mind, we can describe a viewport in terms of a triple (x, y, h), where $h > 0$ is the height above which one is viewing the point on the plane (x, y). This turns the space of viewports into an open three-dimensional halfspace $(x, y, h) \in \mathbb{R}^3, h > 0$.

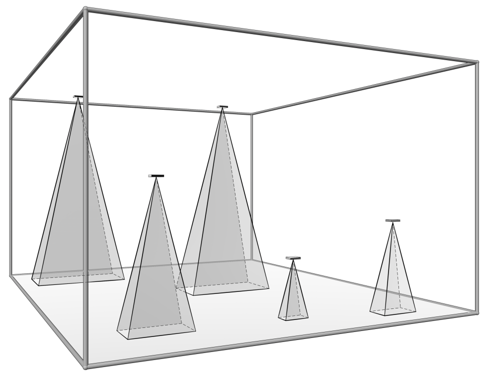

Figure 1: Space of viewports.

Of course, the flat-Earth model used leads to some strange distortions: flying over the terrain towards a pole pushes your viewport closer to the ground. One can do a simple experiment: open your map app with a viewport high enough, above a point somewhere in, say, the Northern hemisphere, and start swiping the view down, i.e., start flying North; the scale bar will grow. We will ignore these inconsistencies in the interpretation of h as literal viewport height and focus instead on the problem of navigation in viewport space.

If you want to move quickly from one viewport to another, how do you do it? Your hand knows better than your head what to do, so natural and intuitive is the interface. Another quick experiment: open a map app on your smartphone if you have one handy, and point your map to, say, this spot 10 miles SW of Shalginsky (a hamlet in Zhanaarka district in Kazakhstan).

Now, navigate in your app to the viewport above the place where you are (the starting point was chosen to increase chances you are not somewhere too close to it).

Traveling between the start and the target physically would be a big deal, crossing a few borders, perhaps some war zones. Moving the viewport is both easy and quick. In some sense, the app gives the impression that any two points are close-at-hand.

The finger metric

The space thus far defined – the *viewport space* $\mathbb{V} := \{(x, y, h), x, y \in \mathbb{R}, h \in \mathbb{R}_{>0}\}$ – is at base level simply a set of points. The additional structure it carries is that of a *metric space* via a *distance function* d. As the tools of navigating the viewport space are zooming in and out, and scrolling in all directions, one can define this distance function somewhat physically, in terms of how much finger movement one needs to get from one viewport to another. Whichever interpretation one adopts – say, the total muscle energy, or ionic flux needed to activate the muscles, or the cortical oxygen

[1]One can also rotate the viewport; this flat circular factor will be ignored, and we will fix the compass throughout.

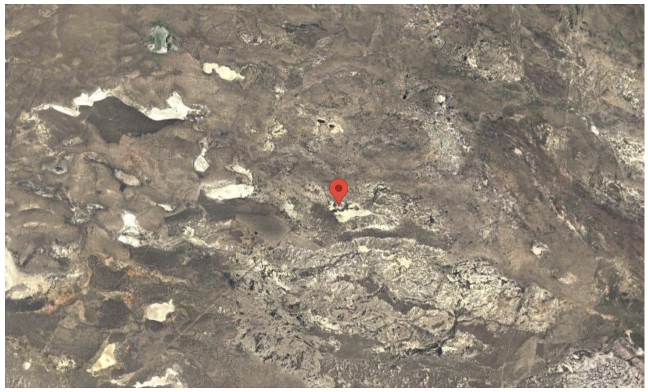

Figure 2: Are you there?

expenditure, one way or another there is a distance function

$$(v_1, v_2) \mapsto d_f(v_1, v_2),$$

which we call *the finger distance* between the viewports $v_1 = (x_1, y_1, h_1)$ and $v_2 = (x_2, y_2, h_2)$.

We cannot speculate too much on the precise shape induced by that metric (how would one measure the amount of muscle energy needed?), but the following should be evident:

- The distance is invariant with respect to horizontal shifts:

$$d_f((x_1, y_1, h_1), (x_2, y_2, h_2)) = d_f((x_1 + a, y_1 + b, h_1), (x_2 + a, y_2 + b, h_2)) :$$

  Indeed, Euclidean motions of the underlying map commute with finger actions.

- The distance is invariant with respect to *dilatations*:

$$d_f((x_1, y_1, h_1), (x_2, y_2, h_2)) = d_f((\lambda x_1, \lambda y_1, \lambda h_1), (\lambda x_2, \lambda y_2, \lambda h_2))$$

  That is, finger motions do not depend on the scale of the map on the screen.

These two observations show that while the finger metric is nearly impossible to quantify experimentally, it possesses a significant group of symmetries; in fact, these symmetries act transitively on the viewport space $\mathbb{V}$.

One more property of the finger metric is important. Namely, the distance is realized by steps of bounded size: as one scrolls or zooms in or out, each of these basis movements changes the viewport in a limited way, and the navigating between viewports that are far apart is done by a composition of these elementary steps.

This last property deserves a brief discussion. In geometry, one often uses the concept of *length space* structure [BBI01], in which the distance between any pair of points can be obtained by finding the *shortest path* between two points. Here a path is a continuous map of the unit interval whose image connects the endpoints. The length of the path is defined by taking a limit of increasingly finer subdivisions of the unit interval domain and summing up the distances between the neighboring points in the subdivision. In length spaces, the distance function can be integrated from local patches. Riemannian manifolds are naturally equipped with the structure of length space. On the other hand, the Euclidean distance between points of a subset of a Euclidean space (like the Earth's surface) is *not* a length space structure, unless the subset is convex.

In our finger metric, we cannot really hope to have a length space, but the palliative of *short steps* is good enough. In fact, an entire research domain of *large-scale geometry* grew out of the desire to view discrete spaces (like lattices or fundamental groups) as being roughly continuous when viewed from afar (see, e.g. [BS07]).

Hyperbolic space

Returning to the symmetries of our finger metric, one can recognize in them (subgroup of) the symmetries of a much more classical space. Recall that the *half-space model* of hyperbolic 3-space is the open half-space $\mathbb{V}$ (same as the viewport space) with distance function characterized by the following properties:

The distance function comes from a Riemannian structure on $\mathbb{V}$: this means that the distance between a pair of points is the minimum of the length over all (piece-wise) smooth paths connecting these points. The length of a parametrized path $\gamma = \gamma(t)$ is given by

$$\int \sqrt{g\left(\frac{d\gamma}{dt}\right)}dt, \tag{1}$$

where g is some Riemannian form, $g = \sum_{k,l} g^{kl} dx_k dx_l$. More importantly, the hyperbolic metric is invariant with respect to the symmetries of the finger metric: the shifts (in fact, arbitrary Euclidean motions) in the x, y plane, and dilatations $(x, y, h) \mapsto (\lambda x, \lambda y, \lambda h)$.

One can immediately derive from these invariance properties, that the Riemannian metric on the half-space model of the hyperbolic space is given by (after, perhaps, some rescalings of the plane, and of the vertical coordinate)

$$\frac{\sqrt{dx^2 + dy^2 + dh^2}}{h}.$$

Of course, the hyperbolic space possesses more symmetries than these, as one has also *inversions*. Recall that inversion in the Euclidean unit (hemi)-sphere $\{x^2 + y^2 + h^2 = 1\}$ takes

$$(x, y, h) \mapsto \frac{(x, y, h)}{x^2 + y^2 + h^2},$$

and is therefore an involution preserving point-wise the unit (hemi-)sphere, and set-wise, the rays through the origin. The translation and dilatation invariances imply that inversions in any (hemi)sphere with center on the $h = 0$ plane is a symmetry of the hyperbolic space. Some elementary geometry shows that an inversion takes vertical lines to Euclidean half-circles contained in 2-plane spanned by the vertical line and the center of the inversion. These half-circles are centered on $h = 0$ and pass through the center of the inversion.

Hyperbolic spaces are far older than iPhones, Google Maps, and viewports, and thus have a more developed nomenclature. For example, the *ideal boundary* of the hyperbolic space is the horizontal plane $\{h = 0\}$, corresponding to the un-ideal state where the phone falls to Earth.

Geodesic lines in hyperbolic space

In any length space, a *geodesic line* is a path such that the length of any sufficiently short subinterval equals the distance between its endpoints. Geodesic lines in a hyperbolic half-space are classical. First, for any two points in $\mathbb{V}$ above the same point on the ideal boundary, the unique geodesic connecting them is the vertical segment, as contemplation of the integral in (1) confirms. Second,

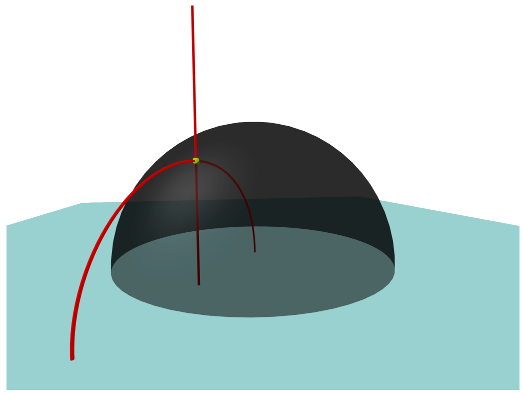

Figure 3: Geodesic lines and inversions in half-space model of hyperbolic space.

for any *Euclidean* half-circle in $\mathbb{V}$ contained in a plane orthogonal to and with the center on the ideal boundary, the inversion in any hemisphere centered at a point of contact between the half-circle and the ideal boundary will take that half-circle to a vertical line. It follows that any arc of such a half-circle is a geodesic between its boundary points.

Of course, such circular geodesic looks curved to us who reside in the Euclidean space used to describe hyperbolic space: for the native dwellers of $\mathbb{V}$, these half-circles are perfectly straight. On the other hand, for them what we consider a straight line at the constant height h would be curving up. The negative curvature of hyperbolic space is responsible for these and many more strange phenomena.

One concludes that for any pair of points in $\mathbb{V}$ with h-values relatively small compared to the Euclidean distance between their projections to the ideal plane, the geodesic trajectory connecting them first shoots up, and then descends back, traversing ballistically in a near semi-circle.

Quasiisometries

One gets the distinct impression that the finger metric and the hyperbolic metric are not exactly the same and yet are qualitatively similar: they are defined on the same set, and share swipe and pinch symmetries. The precise term that describes the relation between these two metrics is *quasiisometry*.

Quasiisometries are, as the name suggests, nearly but not quite isometries – mappings between spaces that preserve the distances *exactly*.

Definition 1. *A transformation $f : X \to Y$ between metric spaces is a* quasiisometry *if for some constants $C > 0, L > 1$*

$$\frac{1}{L} d_X(x, x') - C \leq d_Y(f(x), f(x')) \leq L d_X(x, x') + C. \tag{2}$$

The meaning is straightforward: under the mapping, distances are corrupted only up to a multiplicative factor, after some initial setup cost C is taken into account. Such a definition allows, for example, the inclusion map of the integers into the real number line to be a quasiisometry. The image of a quasiisometry can be thin in the target space (integers can be embedded quasiisometrically into the complex plane too). This is not the case if the range of a map $f : X \to Y$ is a *net* in Y (meaning every point in Y is a uniformly bounded distance away from some point in $f(X)$), in which case the spaces X and Y are called *quasiisometric*.

As is clear from the definition, the notion of quasiisometric spaces is meaningful only at the large scale, where the additive constant C becomes irrelevant. One might wonder if the presence of that $L > 1$ gives too much freedom to distort. And indeed, in Euclidean space, for a slack L arbitrarily close to 1, a path with length $Ld(x, y)$ connecting two points, can stray away from the geodesic connecting x and y arbitrarily far, if $d(x, y)$ is large enough. Amazingly, this does not happen in hyperbolic spaces (and their quasiisometric kin), as we will see below, with implications for the finger metric.

We want to deploy the notion of quasiisometry in our situation of metric spaces sharing the set $\mathbb{V}$ where they are defined,[2] just with different metrics structures: one, the standard hyperbolic metric on the half-space, and the other the finger metric.

Proposition 2. *These two metric spaces are quasi-isometric.*

The proof works for any pair of metrics where distances are realized by paths of bounded step sizes, and which share a point-transitive group of isometries. Indeed, the unit ball in the finger metric centered at any point is squeezed between two balls in the standard hyperbolic metric of radii L and $1/L$, for some L. Existence of a transformation preserving both metrics and taking that point to any other point then makes it true uniformly. Hence, with potential correction for the first and last step, the lengths of the paths realizing either metric are bounded from above and below by the inequalities (2).

Quasi-implications

What can the notion of quasiisometry of two metrics on a space deliver? Clearly, at best some statements about large-scale structure: the quasiisometry of any lattice to the vector space it spans is an intuitive example of how much slack the notion affords.

One immediate corollary of the definition is that any path realizing the finger distance between a couple of points is necessarily a L-*quasi-geodesic* in the hyperbolic metric: it means that the finger distance between any pair of points along the path is bounded by L times the hyperbolic distance between the points (with a uniform additive slack): one can just take the hyperbolic geodesic connecting these points and split it into the small steps of the finger metric. Note that the quasi-geodesic need not be a *path* – a perfect fit to the sequences of short steps in the viewport spaces.

One of the most salient properties of quasi-geodesic paths in hyperbolic space is that (unlike Euclidean spaces) they sheepishly follow the actual hyperbolic geodesic lines. This remarkable fact is known as the *Morse lemma* and is at the foundation of the beautiful theory of *Gromov hyperbolic spaces* (see e.g., [BBI01] for a survey).

A length space is called δ-hyperbolic if for any three points A, B, C, the geodesic path connecting A and B lies within the union of δ-tubes around the geodesic trajectories connecting AC and BC. This property (*the thinness of triangles*) is quite easy to deduce in the standard hyperbolic plane. Indeed, there, the sum of the angles of any triangle is π minus the area of that triangle. In particular, any triangle, and hence any circle inscribed into the triangle has an area of at most π which gives 2-hyperbolicity for the plane (the areas of the circles in the hyperbolic plane grow faster than the areas of Euclidean circles). In the standard hyperbolic space of any dimension, we have the same property, as any triangle lies in a hyperbolic plane.

Hence, standard hyperbolic spaces are δ-hyperbolic, for some $\delta \leq 2$. Passing to δ-hyperbolic spaces allows one to vastly generalize the scope of applications, and at the same time to make the proof of the Morse lemma [BH99] rather elementary (if very clever)[3].

[2] This would render redundant the condition on the range of f to be a net.

[3] and much more transparent than the original approach taken by Marston Morse in his thesis [Mor21].

Proposition 3 (Morse lemma)**.** *In a δ-hyperbolic space, the Hausdorff distance between a quasi-geodesic (with additive and multiplicative slacks C and L) and a geodesic segment sharing the same endpoints is bounded by a constant depending only on C, L and δ and not on the length of the geodesic.*

This means that the geodesic of our finger metric - which is a quasi-geodesic in the hyperbolic metric on the space of viewports $\mathbb{V}$, - sits, by the Morse lemma, within a tube of uniformly bounded width[4] around the geodesic of the hyperbolic metric which we know precisely.

Geodesic pinch-swipe

The Morse lemma greatly simplifies an analysis of the optimal navigation of the viewport space. While an exact computation of finger distance or finger geodesic paths between a pair of viewports is a meaningless exercise (depending too much on individual neurophysiology, and the parameters of the mobile device one is using), the Morse lemma tells us that unless the path is exponentially worse than the optimal one in the finger metric, it will always be at a uniformly bounded distance from the optimal path in the *hyperbolic* metric on the viewport space.

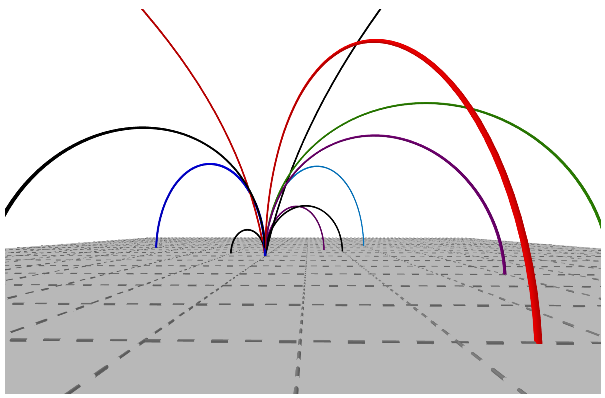

Figure 4: All geodesics of the finger metric starting at a viewport close to the ideal boundary need to follow closely the ballistic trajectories of the geodesics of the hyperbolic metric.

In other words, any reasonable navigation in the viewport space needs to be what we do intuitively: zoom the original view out till the desired point on the map is within the viewport (or almost there), then zoom back in, centering on the target point. The resulting trajectory is necessarily a quasi-geodesic in the hyperbolic metric.[5]

Concluding Remarks

The net cost of the quasigeodesic paths cannot be improved (up to a multiplicative factor), not just in the path metric situation, but for any input method. Indeed, the volume of the space of viewports (say, as measured in finger unit-ball volumes) is a bounded multiple of the number of non-overlapping

[4]The latest best bound on this width of the tube, in terms of the quasi-isometry constant L, thinness parameter δ and the step sizes S of the quasiisometric path, is $92S^2(L+\delta)$ [GS08].

[5]The beauty of two far-off points on Earth being connected high-above by a pair of pinch-zooms recalls Donne's image of the compass connecting distant sublunary lovers:

"If they be two, they are two so As stiff twin compasses are two" – A Valediction Forbidding Mourning viewports of maximal resolution, $N = S/A$), where S is the total area of the underlying plane map of the terra cognita, and A is the area of the smallest surface patch that is renderable as a viewport. (The total Earth area, $\approx 5 \cdot 10^{14}\mathrm{m}^2$ is a lower bound on S; a typical device gives $A \approx 10^2\mathrm{m}^2$, giving $S/A \geq 5 \cdot 10^{12}$ – quite a lot of space to cover.)

In the half-space model, $S^{1/2}$ is the order of magnitude of the maximal (Euclidean) distance between the projections of pair of points to the ideal boundary, and $A^{-1/2}$ is the height h of the viewport. One can easily see that the largest hyperbolic distance between the viewports is therefore of order $\log(A/S) = \log(N)$, the logarithm of the number of viewport of maximal resolution. Merely to *encode* the position of a viewport (a task implicitly contained in that of navigation to the targeted viewport) one would need to enter that many bits, as the viewport identifier, be it the latitude/longitude pair, or what3names id, etc. Thus one would need to expend the same (up to a factor) ionic flux or muscle effort as the navigation of the viewport space would require. So, the hyperbolic navigation routine is optimal up to a factor, and, of course, far more intuitive than typing in the coordinates of a spot.

The idea that hyperbolic geometry is well-suited for navigating various spaces is well established among computer scientists and engineers. It has been suggested as a good template to embed semantic networks [LR13, AC20], or to model the Internet [CDE+12, JLBB11]. Yet, it seems that the fact that a widely used mobile application has hyperbolic geometry as its foundation has not been registered in the literature. There is something deeply satisfying in understanding *why* using this applications is so convenient: the *feeling* that the procedure is efficient is based on the underlying geometry which we perceive, perhaps, subconsciously.

References

[AC20] Maria Astefanoaei and Nicolas Collignon. Hyperbolic Embeddings for Music Taxonomy. preprint, PsyArXiv, October 2020.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BH99] Martin R. Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[BS07] Sergei Buyalo and Viktor Schroeder. *Elements of asymptotic geometry*. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich, 2007.

[CDE+12] Victor Chepoi, Feodor F. Dragan, Bertrand Estellon, Michel Habib, Yann Vaxès, and Yang Xiang. Additive Spanners and Distance and Routing Labeling Schemes for Hyperbolic Graphs. *Algorithmica*, 62(3-4):713–732, April 2012.

[GS08] Sébastien Gouëzel and Vladimir Shchur. A corrected quantitative version of the morse lemma. *Journal of functional analysis.*, 277(4), 2019-08.

[JLBB11] Edmond Jonckheere, Mingji Lou, Francis Bonahon, and Yuliy Baryshnikov. Euclidean versus hyperbolic congestion in idealized versus experimental networks. *Internet Mathematics*, 7(1):1–27, 2011.

[LR13] Gregory Leibon and Daniel N. Rockmore. Orienteering in Knowledge Spaces: The Hyperbolic Geometry of Wikipedia Mathematics. *PLoS ONE*, 8(7):e67508, July 2013.

[Mor21] Harold Marston Morse. A one-to-one representation of geodesics on a surface of negative curvature. *American Journal of Mathematics*, 43(1):33–51, 1921.