Math-801

A. Assadi

Introduction to Parallelism on Surfaces

In Euclidean plane geometry, the notion of parallel lines is inspired and follows the intuition that we gain from visual perception of our environment. The procedure could follow a scenario like this. First, the space is perceived not as an empty entity void of any physical objects or forces; rather, the notion of space is perceived or formulated as a convenient abstraction of a container in which objects are situated or have movement, growth, deterioration etc. This view of space is very much like taking the coordinates in analytic geometry. When we introduce a *metric* to measure distances, lengths and angles, we are led to distinguishing objects and notions that stand out from the rest by virtue of having unique or rare properties. In the parlance of vision science, such distinguishing properties or entities are called *salient features*, and the extent that they impress upon our senses (and consequent higher level brain processes) is referred to as *saliency*. Among all paths joining two points P and Q in the plane, the path with shortest distance stands out both perceptually in terms of its frequency of occurrence in the environment and its physical measurements that imply practical consequences in terms of saving animal effort when work and expenditure of energy occupies attention. Straight line is therefore unique by virtue of a salient feature that is a consequence of our visual perception of the environment and the laws of physics as they are experienced through our senses. The notion of parallel lines follows course once the class of straight lines are singled out by observers by virtue of their salient property of being the unique path in the space joining two points and having the shortest length. Parallel lines also occur quite naturally in the environment, especially in the vertical direction due to the effects of gravity, and the empirical physical Principle of Least Action which affects any kind of dynamic phenomena, from growth of tree trunks in shaded sites, free fall of rocks, to formation of icicles and stalagmites. On the other hand, when randomly drawing a pair of straight lines, there are far fewer pairs of parallel lines that are drawn naturally than pairs of intersecting lines. Indeed, human effort is needed to maintain parallelism of two lines drawn on even a plane surface. The saliency of parallelism among such random pairs of lines in the plane singles out the germ of definition of a line being parallel to another, or more generally, a direction being parallel to another direction. The relationship of parallelism with the angle of intersection of pairs of lines is also apparent based on saliency of parallelism versus the abundance of intersecting pairs with all kinds of random measures for the angle of incidence. Now take a pair of parallel line Π_1 and Π_2, and intersect them with another straight line L, that almost always would intersect them, and let L intersect also a number of other pairs that are NOT parallel and have randomly varying angular measurements when observing their angle of incidence. A typical pair of intersecting lines Λ_1 and Λ_2 intersects L in two quite different angles of incidence for each one of Λ_1 and Λ_2, and this "*having non-equal pairs of angles of incidence*" is the rule rather than the exception! Indeed, the property that is observed when L intersects a pair of parallel line Π_1 and Π_2 is a rare event, and stands out compared to the randomly varying, irregular patterns of the angular sizes for L intersecting typical pairs such as Λ_1 and Λ_2. The fact that angles of incidence of a line L intersecting a parallel pairs of lines Π_1 and Π_2 are equal becomes a significant observation when we encounter it in numerous regular figures, such as rectangles, parallelograms and other regular polygons with even numbers of sides. The latter observation is a simple theorem in Euclidean geometry, and its simplicity is the key to its power when we generalize the notion of parallelism to spaces and geometries that are not necessarily Euclidean.

Now we rearrange the same collection of observations into the following form. Fix a direction in the plane, and consider a straight line L with incidence angle θ_1 with a line Π_1, say drawn from a point P on the line L in the given direction. Then select the line Π_2 drawn from another point Q on L in the same direction, and call θ_2 the angle of incidence of Π_2 intersecting L. Then the angles of incidence are the same, and conversely, when we draw a line from the point Q on L such that it forms the angle θ_1 with L, then the new line will be parallel to Π_1, in other words, it will be in the same direction as Π_1 only drawn from the point Q. The application of this formulation is far greater than one might expect. Just think of all the theorems in plane geometry that tell us when two figures (say two triangles) are congruent or similar or are related in any of the numerous ways they could be! The empirical validation of such statements and even the rigorous logical proof almost always involves translating lines and other geometric objects parallel to a given direction, and the property that is often taken for granted due to its very natural perception is that such a translation does not change the metric relationships between pairs of points, pairs of angles from the first figure and its translated copy.

How should we generalize the notion of Euclidean parallelism of two straight lines to surfaces and other geometric non-Euclidean structures? First, we must settle the issue of replacing straight lines with the appropriate curves on a surface S according to a property that would be "*salient*" among all arbitrarily or randomly drawn curves that connect two points P and Q. It turns out that the crucial property of minimizing the length of the curve joining two points is the right one when we restrict attention to pairs of points that are not "too far apart". The safest way to pronounce this is the following: Given a point P on S and any other point Q on S that is sufficiently near P, there is a unique curve joining P to Q that has the smallest length among all paths between the two points. Such a distinguished curve is called a geodesic, and its existence and uniqueness is a direct consequence of the fundamental theorem that describes the solution of quite general ordinary differential equations. Therefore, a smooth surface possesses a class of distinguished curves that connect pairs of points with the shortest length[1].

Now, let us take the statement about presence of geodesics as curves that minimize distance as a fact and move on to explore other properties of a geodesic of a more physical nature. One such property comes from observation of the trajectories of particles moving in the plane and their Newtonian kinematics. Motion of a particle along curves $\alpha(t)$ creates two vector fields along $\alpha(t)$; namely, the velocity $\dot{\alpha}(t) = \dfrac{d\alpha}{dt}$ and the acceleration $\ddot{\alpha}(t) = \dfrac{d\dot{\alpha}}{dt}$. In Euclidean spaces, any curve $\alpha(t)$ could be always re-parameterized by arc length, so that it could be always regarded as the trajectory of a particle moving along with unit speed; that

[1] This hints also to an intrinsic way (i.e. internal to the geometric structure of the surface S and with no reference to the world outside of the surface S) to define a standard for measurement of distances between points on a surface. We can use geodesics to define distance between points on the surface S using the length of the single geodesic joining the pair of points in question.

is $\|\dot{\alpha}(t)\|=1$ and the velocity satisfies $\left\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right\rangle = 1$. Differentiating the latter, we have: $\frac{d}{dt}\left\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right\rangle = 0$, so $\left\langle \frac{d^2\alpha}{dt^2}, \frac{d\alpha}{dt} \right\rangle + \left\langle \frac{d\alpha}{dt}, \frac{d^2\alpha}{dt^2} \right\rangle = 0$. This means that the acceleration of the motion is always perpendicular to the velocity vector. For a straight line passing through a point P, the explicit formula $\alpha(t) = P + t\vec{v}$ shows that the velocity vector is constant $\frac{d\alpha}{dt} = \vec{v}$ and $\frac{d^2\alpha}{dt^2} = 0$. On a curved path, a particle experiences a velocity vector that is not constant, and the changes in its velocity vector's directionality are measured by the acceleration which exerts force on a massive particle. If the curve $\alpha(t)$ lies on a surface that is curved unlike the plane, the velocity vector field cannot be assumed to be given by a constant vector, so the general situation for a curve on a general surface could only require that the acceleration vector field be "as simple" as it could be (from the viewpoint of the particles traveling along curves lying on the surface S situated in the 3-dimensional space) without being zero. The acceleration vector $\frac{d^2\alpha}{dt^2}$ has a decomposition into two types of components: the first type is the projection along the surface, i.e. on the tangent space $T_{\alpha(t)}S$ and the second is the projection in the direction orthogonal to the surface, which means on the subspace spanned by the normal vectors to the surface. When $S \subset \mathbb{R}^3$, there is only one basis vector for the space of normal vectors to the surface at one point. From the viewpoint of the particle motion on the surface, the simplest curves are those that have no acceleration along the surface, and the only component of the acceleration that could be non-zero is the one along the normal to the surface. It turns out that the curves such that $\frac{d^2\alpha}{dt^2} \perp T_{\alpha(t)}S$ automatically must be constant-speed, or equivalently, $\left\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right\rangle =$ constant. This follows from retracing the Leibniz rule for differentiation of the inner product: $0 = \left\langle \frac{d^2\alpha}{dt^2}, \frac{d\alpha}{dt} \right\rangle = \frac{1}{2}\frac{d}{dt}\left\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right\rangle$, so $\left\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right\rangle$ must be constant. It is indeed a theorem in differential geometry that the two definitions are equivalent for n-dimensional hypersurfaces of $\mathbb{R}^{n+1}$ (level surfaces of regular values of smooth maps $\mathbb{R}^{n+1} \to \mathbb{R}$).

At this point, we have generalized the properties of straight lines in Euclidean geometry to the point that it is sufficient to formulate the notion of parallelism and "parallel translation" on a surface, or more generally, on a hypersurface that is a level set for a regular value of a smooth function $\mathbb{R}^{n+1} \to \mathbb{R}$. Our first observation is that the process of translation a figure, for example a directed line segment AB parallel to a given direction, generates a *vector field* along the curve that is traced by the point A. This vector field is called parallel in the plane containing the curve, and the process of generating such a parallel vector field is called *parallel transport*. Clearly the notion of parallel transport includes the ingredients for defining when two tangent vectors at two points P and Q are parallel. Let $\alpha(t)$ be a geodesic on S joining two points P and Q. Further, suppose a vector field V is given along the curve $\alpha(t)$, and we wish to investigate when this vector field could be called parallel. It would be quite reasonable to anticipate that if $V(\alpha(t)) = \dfrac{d\alpha}{dt}$ is the velocity vector field of the geodesic $\alpha(t)$, then it is sensible to call V a parallel vector field along $\alpha(t)$. Next, for any vector field $V(\alpha(t))$ along $\alpha(t)$, we can measure the infinitesimal variation of $V(\alpha(t))$ as a tangent vector to S and compare the result with other tangent vector fields. This process follows the method introduced above for testing if a curve is a geodesic by requiring that the acceleration vector field is only along the normal direction. Here, $\dfrac{d^2\alpha}{dt^2}$ is the result of taking the "directional derivative" of $V(\alpha(t)) = \dfrac{d\alpha}{dt}$ along $\alpha(t)$. Thus, if $\dfrac{d}{dt}V(\alpha(t))$ has only a normal component and its tangential component vanishes, then it has no variation along $\alpha(t)$ from the view point of the surface S. This observation indeed suggests that the curve $\alpha(t)$ need not be a geodesic necessarily for our condition of parallelism to be defined. In other words, we simply take the "*directional derivative*" of $V(\alpha(t))$ in the direction of the tangent vector field to $\alpha(t)$. We call $V(\alpha(t))$ parallel if its "derivative" has no tangential component; that is, the only nonzero components are normal to S. Later, we make these ideas into analytic formulas and explore them further using symbolic computation. See HW #6 for concrete examples and more details.