the new contoured world: minkowski's construction of spacetime

Paul A. Obrecht

December 3, 2001

All the empiricists must have laughed when Kant defined space and time as “two pure forms of sensible intuition”[1] (67). He claimed space to be “nothing but the form of all appearances of outer sense. It is the subjective condition of sensibility…” (71). Time, on the other hand, was “nothing but the form of inner sense, that is, of the intuition of ourselves and of our inner state” (77). For centuries before, extension in space had been one of the fundamental properties of objects, and everything (save perhaps the Divine Architect himself) was said to exist in the sequential, directed medium of time. Kant turned all that on its head to say that space and time, at least as we experience them, are in fact properties of the mind that condition our mode of perception. He erected an insurmountable wall between our internal model of the universe—three Euclidean space dimensions and one progressive entity called time—and the things in themselves. This was the first hint that space and time as such might be nothing but useful fictions which nonetheless appear deceptively real.

Fast forward a century and a quarter, and envision yourself in the audience at Hermann Minkowski’s 1908 lecture demonstrating “how it might be possible, setting out from the accepted mechanics of the present day, along a purely mathematical line of thought, to arrive at changed ideas of space and time”[2] (75). His claim is that a purely mathematical derivation of the relationship between space and time, as described by Lorentz and Einstein, had been possible since the first speculations about light’s constant velocity for all observers. Any astute mathematician, he says, could have derived the same results, and the overarching geometrical model, given only the constant velocity of light and the two-fold invariance of Newtonian mechanics as axioms. In fact, Minkowski goes so far to say that any mathematician since Newton could even have predicted the constant velocity of light, long before empirical evidence suggested it, at the same time that he realized that “natural phenomena do not possess an invariance with the group G_∞, but rather with a group G_c, c being finite and determinate, but in ordinary units of measure, *extremely great*” (79).

The elegance with which Minkowski is able to coax the truths of relativity from a geometrical figure is astounding, and in the end it strongly suggests that the underlying structure of the Universe must bear some resemblance to the manifold he proposes. In the end, we must keep in mind Kant’s suspicions about space and time, but that should not prevent us from exploring the implications

[1] I. Kant (1781), *Critique of Pure Reason*, translated by N. K. Smith (New York: Macmillan, 1929).

[2] H. Minkowski (1908), “Space and Time.” In A. Einstein *et al.*, *The Principle of Relativity*, translated by W. Perrett and G. B. Jeffery (New York: Dover, 1952).

that Minkowski's surface so convincingly brings before our eyes. To that end, we embark on an exploration of this unique surface, attempting to present a clearer image of it to the mind's eye, to explicitly derive some of the implications present just beneath the surface of its formula, and to demonstrate the power it has to generalize Einstein's revelation into a broader context. We begin with the two Newtonian invariances.

The first invariance corresponds to a geometrical principle—the arbitrary rotation of the coordinate space axes, x, y, z, provided that the quantity $x^2+y^2+z^2$ remains constant. This condition explicitly guarantees that the space axes remain, under the given rotation, perpendicular to each other. Though the tools of measurement may be changed arbitrarily, what is measured must clearly remain invariant in any particular transformation of coordinates. Thus distances reckoned in any coordinate system must remain equal, and as long as the orthogonality of the coordinate axes is maintained, $d^2 = x^2 + y^2 + z^2$ remains constant. But an implicit consequence of this condition amounts to an assumption about the nature of space itself; by requiring that distance remain invariant through any "arbitrary change of position" (75), Minkowski necessarily adheres to the notion that every piece of space is identical, that is, that space is uniform and Euclidean in every direction. This assumption might seem quite reasonable, but Kant's *Critique of Pure Reason* has warned that Euclidean space may be merely a principle of the mind, and Gauss' *General Investigations of Curved Surfaces* (1827) and Riemann's *On the Hypotheses which Lie at the Bases of Geometry* (1854) presciently confirmed that Euclidean space is merely a trivial case of all possible spaces. These revolutions in geometrical thought would later allow Einstein to base his general theory of relativity on the notion of curved space, in which the measure of distance can vary with position.

The second invariance corresponds to a physical principle—that of Galilean relativity—in which uniform motion is said to be utterly indistinguishable from rest. This principle, in Minkowski's formulation, is interpreted geometrically to indicate a change in the orientation of the time axis with respect to the space axes, accomplished by replacing the space axes x, y, z with $x-\alpha t$, $y-\beta t$, $z-\gamma t$. Thus any observer in uniform motion can consider his time axis to be oriented parallel to the direction of his travels through space, rendering him stationary in his own frame of reference. Such a geometrical interpretation, though, requires the construction of a four-dimensional manifold with the four axes x, y, z, t. In this new manifold, later dubbed "spacetime," the three space axes must always remain perpendicular to each other (as a consequence of the first invariance), but the time axis can have any orientation to the others, given certain constraints dictated by the parameter c (constraints absent from Newton's notions of space and time). The time axis, nonetheless, remains an independent variable, and thus it uniquely determines a fourth dimension, independent of the other three.

Minkowski summarily dismisses the problems inherent in making the leap to four dimensions, saying, "the somewhat greater abstraction associated with the number four is for the mathematician no infliction" (76). We should not be so quick to agree. If geometry is, at bottom, a visual endeavor, then venturing out into realms in which we can argue only by analogy to our sense experience in three dimensions requires, to some degree or another, a leap of faith. We are forced to accept Minkowski's extraction of the underlying algebraic relationships from two- or three-dimensional geometry, and his application of the same to manifolds of higher dimension by adding variables. Algebraically, this operation presents no real problem, but in doing so we only conceal the question about the possibility of generalizing our space notions. Reading Minkowski's manifold as an appropriate mathematical model, one which manifests the relationships under consideration, serves as a useful tool, but ascribing to it real physical existence is a much greater leap. His time- and space-equating "mystic formula" indicates that his intent is the latter.

Philosophical questions aside for the moment, we return now to the mathematical details associated with the number four. Any constant function of four independent variables, of course, defines a four-dimensional surface. (For instance, the function $t^2 + x^2 + y^2 + z^2 = 1$ represents a four-dimensional surface analogous to a unit sphere, while $t^2 = 1$ [x, y, z constant] defines a plane-analogue.) But only one such surface possesses the required invariances, and Minkowski invites consideration of that particular surface, $c^2t^2 - x^2 - y^2 - z^2 = 1$, because he knows from prior investigation what will result. Since hyper-dimensional arguments must be presented in some degree by analogy to our two- or three-dimensional imagination, he demonstrates it only for a planar cross-section of the surface in the x-t plane. The argument essentially runs like this: Any given hyperbola can be described by an infinite number of conjugate diameter pairs, and if the ratio of units on those conjugate diameter pairs remains always constant (in this case [unit d]/[unit t] $= c$ for all choices of x- and t-axes), the planar cross-section is always given, regardless of the choice of axes, in the form

$$c^2t^2 - x^2 = 1.$$

This result is just as true of the four-dimensional case, which is easily demonstrated by the application of an algebraic generalization. We know from the restrictions of the first invariance that the measure of distance in any two coordinate systems must remain equal. Thus we set

$$\begin{aligned}
x^2 + y^2 + z^2 &= (x - \alpha t)^2 + (y - \beta t)^2 + (z - \gamma t)^2 \\
x^2 + y^2 + z^2 &= (x^2 - 2x\alpha t + \alpha^2 t^2) + (y^2 - 2y\beta t + \beta^2 t^2) + (z^2 - 2z\gamma t + \gamma^2 t^2) \\
0 &= -2x\alpha t + \alpha^2 t^2 - 2y\beta t + \beta^2 t^2 - 2z\gamma t + \gamma^2 t^2 \\
2x\alpha t + 2y\beta t + 2z\gamma t &= \alpha^2 t^2 + \beta^2 t^2 + \gamma^2 t^2 \\
2t(\alpha x + \beta y + \gamma z) &= \alpha^2 t^2 + \beta^2 t^2 + \gamma^2 t^2. \qquad (1)
\end{aligned}$$

We now make the substitutions representing the inclined time axis into the expression for Minkowski's surface and simplify.

$$\begin{aligned}
c^2t^2 - x^2 - y^2 - z^2 &= 1 \\
c^2t^2 - (x - \alpha t)^2 - (y - \beta t)^2 - (z - \gamma t)^2 &= 1 \\
c^2t^2 - (x^2 - 2x\alpha t + \alpha^2 t^2) - (y^2 - 2y\beta t + \beta^2 t^2) - (z^2 - 2z\gamma t + \gamma^2 t^2) &= 1 \\
c^2t^2 - x^2 - y^2 - z^2 + 2t(\alpha x + \beta y + \gamma z) - (\alpha^2 t^2 + \beta^2 t^2 + \gamma^2 t^2) &= 1.
\end{aligned}$$

Substituting the value from Equation (**??**) for the $2t(\alpha x + \beta y + \gamma z)$ term above gives

$$c^2t^2 - x^2 - y^2 - z^2 + (\alpha^2 t^2 + \beta^2 t^2 + \gamma^2 t^2) - (\alpha^2 t^2 + \beta^2 t^2 + \gamma^2 t^2) = 1.$$

The last two terms on the left hand side are identically equal for any values α, β, γ, which leaves the original expression for the surface:

$$c^2t^2 - x^2 - y^2 - z^2 = 1.$$

There is an additional complication in the four-dimensional case that is not explicit in the algebraic derivation. Minkowski's Figure 1 elucidates that an inclination of the t-axis down from the vertical toward the asymptote is mirrored by an inclination of the x-axis up from the horizontal toward the asymptote. In the four-dimensional case, the x-axis still inclines toward the asymptote, but the orthogonality condition requires that the y- and z-axes also incline (and/or rotate) in order to remain mutually perpendicular. Thus Minkowski's four-dimensional spacetime consists of three Euclidean space axes hinged to an arbitrarily-oriented time axis so that the space axes retain a given orientation to each other while the time axis is free to swivel about in relation to them given, again, certain constraints determined by the parameter c. These constraints amount to a condition that the time axis must lie within the four-cone analogous to the asymptotes in a plane-hyperbola. Parenthetically, it is important to note that Minkowski's derivation does not predict a definite value for c, since his derivation is valid for any positive finite value c. Only in correlating the mathematics to the phenomena does the physical interpretation of the parameter c takes on a specific value: the speed of light, or in other terms, to avoid unnecessary reliance on empirical data, the "ratio of the electromagnetic to the electrostatic unit of electricity" (79).

Though not apparent on the surface, and despite Minkowski's understatement of its significance, his construction amounts to a revolution in the notion of distance. In order to understand why, note Minkowski's "fundamental axiom" (80): *The substance at any world-point may always, with the appropriate determination of space and time, be looked upon as at rest.* This axiom is a direct logical consequence of the four-hyperboloid's invariance, and its significance can be given algebraically.

$$
\begin{aligned}
c^2dt^2 - dx^2 - dy^2 - dz^2 &> 0 \\
c^2dt^2 &> dx^2 + dy^2 + dz^2 \\
c\,dt &> \sqrt{dx^2 + dy^2 + dz^2} \\
c &> \sqrt{(dx/dt)^2 + (dy/dt)^2 + (dz/dt)^2} \\
c &> v.
\end{aligned}
$$

Any velocity must necessarily be less than the speed of light, as Minkowski notes. This result is independent confirmation of Einstein's, from a purely geometric framework. Yet Einstein's derivation of this relationship implicitly treats the space dimensions differently from the time dimension: Distance is still measured in one realm, time in another, and velocity is merely a relation between the two. Minkowski's construction of spacetime, on the other hand, forces us to broaden our notion of distance, for his intention in uniting space and time is to present a static whole, in which they appear different only because of our limited imagination. By untethering our imagination, he would say, the reality suggested by our sense perceptions will "fade away into mere shadows" of something far greater.

Consider then a new definition of distance through the four-dimensional manifold, a function of the four variables x, y, z, t. Minkowski calls this function τ when it is time-like, and gives the length of infinitesimal "distance" $d\tau$ as (80)

$$d\tau = \frac{1}{c}\sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2}.$$

The negative signs are the critical elements of the new notion, for distance has never been reckoned like this. Euclidean distance in three dimensions is given by $ds = \sqrt{dx^2 + dy^2 + dz^2}$; its analogue in four dimensions would include a dt^2 term, but all the terms would remain positive. In the above expression for distance through spacetime, on the other hand, the Euclidean space distance is *subtracted* from the distance along the time axis, suggesting that the primary motion of any substantial world point is *through time*, and that motion through space is merely a deviation from this primary motion.

By contrast, we can easily imagine Minkowski's surface residing in a four-dimensional manifold in which the Euclidean notion of distance holds. In this case, the equation for the surface would remain the same, but distance would then be reckoned by the appropriate version of the Pythagorean theorem. Since all points on the surface must satisfy the formula $c^2t^2 - x^2 - y^2 - z^2 = 1$, the time coordinate can be rewritten as a function of the other three:

$$t = \frac{\sqrt{x^2 + y^2 + z^2 + 1}}{c}.$$

Any point on the surface can thus be given by the general coordinates

$$\left(x, y, z, \frac{\sqrt{x^2 + y^2 + z^2 + 1}}{c}\right),$$

and the distance from the origin to the point is given by (substituting the above value for t)

$$d = \sqrt{x^2 + y^2 + z^2 + t^2} = \sqrt{\frac{(c^2 + 1)(x^2 + y^2 + z^2) + 1}{c^2}},$$

which clearly varies as the coordinates x, y, z vary. Minkowski's surface considered in this way to reside in Euclidean space retains the two required invariances but does not retain the constant ratio of units. As a result it remains inadequate for the purposes of expressing the truths of relativity.

On the other hand, Minkowski's new definition of distance has the counterintuitive property that the distance from the origin to any point on the surface remains always the same. This fact is a trivial algebraic consequence of defining distance with the same expression as defines the surface. For example, the Minkowskian distance from the origin to the points on the surface (1, 1, 1, [$2/c$]) and (4, 4, 4, [$7/c$]) is

$$\sqrt{c^2\left(\frac{2}{c}\right)^2 - 3(1^2)} = \sqrt{c^2\left(\frac{7}{c}\right)^2 - 3(4^2)} = 1.$$

The consequences are extraordinary. Any world point in uniform motion passing through the origin at $t = 0$ will necessarily be found, after one unit of its proper time, somewhere on the surface. But we can even generalize this result to non-uniform motion. Turning the distance expression into an expression of the velocity through spacetime by referring both sides to a time-element dt, gives a form of the Lorentz time transformation—as expressed by Einstein on page 49—valid for noninertial (accelerated) reference frames.

$$\begin{aligned}
d\tau &= \frac{1}{c}\sqrt{c^2dt^2 - dx^2 - dy^2 - dz^2} \\
d\tau^2 &= \frac{1}{c^2}(c^2dt^2 - dx^2 - dy^2 - dz^2) \\
\frac{1}{dt^2}d\tau^2 &= \frac{1}{dt^2}\frac{1}{c^2}(c^2dt^2 - dx^2 - dy^2 - dz^2) \\
\left(\frac{d\tau}{dt}\right)^2 &= \frac{1}{c^2}\left[c^2\left(\frac{dt}{dt}\right)^2 - \left(\frac{dx}{dt}\right)^2 - \left(\frac{dy}{dt}\right)^2 - \left(\frac{dz}{dt}\right)^2\right] \\
\left(\frac{d\tau}{dt}\right)^2 &= 1 - \frac{1}{c^2}\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right]
\end{aligned}$$

Notating velocity through space alone by $v = \sqrt{(dx/dt)^2 + (dy/dt)^2 + (dz/dt)^2}$, we have

$$\begin{aligned}
\left(\frac{d\tau}{dt}\right)^2 &= 1 - \frac{v^2}{c^2} \\
\frac{d\tau}{dt} &= \sqrt{1 - \frac{v^2}{c^2}} \\
d\tau &= dt\sqrt{1 - \frac{v^2}{c^2}}
\end{aligned}$$

Since v is a vector composition of infinitesimal velocities, the above expression consists of nothing but infinitesimal quantities and the constant parameter c. Thus the Lorentz time transformation simply falls out of the expression for distance, but is now generalized to infinitesimal quantities that allow for integration over a curved world line, and thus over changing velocities. The other transformations could be similarly derived. Einstein was unable to accomplish this. Thinking about things in terms of clocks, rods, and various coordinate systems moving with uniform velocities enabled Einstein to develop the special theory of relativity, the importance of which cannot be overestimated. But the power of Minkowski's geometrical construct enables the generalization of the same relationships to noninertial reference frames, and this new-found power underscores its coherence and validity.

Now, it was shown above that $d\tau$ is a line-element along the world line of a substantial point, and thus it represents an infinitesimal distance through spacetime. But Minkowski notes that $d\tau$ also represents an increment of proper time for the object traversing that world line (85). Thus the proper time of a substantial world point is nothing but the distance it traverses through spacetime. One strange consequence of this equality of distance and proper time manifests itself in the propagation of light. The distance traversed by a ray of light during a time increment is given by its velocity times the time increment, or $c\,dt$. But the distance traversed by the same ray of light is also given by $\sqrt{dx^2 + dy^2 + dz^2}$. Setting the two equal, squaring, and collecting terms on one side gives $c^2dt^2 - dx^2 - dy^2 - dz^2 = 0$. Substituting this into Minkowski's expression for distance clearly indicates that *light experiences no passage of proper time.* Further, the above equivalence between proper time and distance suggests, contrary to all intuition, that a light ray does not traverse any distance. Only manifolds characterized by distance formulas containing opposite signs possess the property that two non-coincident points can be separated by a distance of 0.

Thus any object whose world line is always time-like (as it must be for anything with mass) experiences the passage of proper time. Light-lines, then, which lie on the four-cone between time-like and space-like regions of spacetime, form the boundary case because they experience no passage of proper time. For space-like world lines, $x^2 + y^2 + z^2 > c^2t^2$, so the quantity under the radical is negative. The results are equal in magnitude to their corresponding time-like distances, but space-like distances are always imaginary.

The entire continuum of interplay between space and time, both possible in this world and impossible, is thus represented in Minkowski's spacetime manifold. He continues in the final sections of the paper to deduce various other physical laws and phenomena from the geometry of the surface under investigation, but to address them here would take us too far afield. It suffices to say that, as Apollonius' cone-upon-a-cone defines the framework in which the relationships between the various conic sections are most elegantly displayed, so Minkowski's manifold defines that in which the relationships between space and time find “their most perfect expression” (76).

Or nearly.

If Minkowski was testing the doorknob, Einstein later kicked open the door to a new vision of geometry. An impulse toward closure demands that we take a brief sneak peak ahead to Einstein’s 1916 paper in order to see his response to Minkowski. Einstein must have been highly impressed by the spacetime construct, for he incorporated a four-dimensional manifold into his general theory of relativity, but Einstein also read Reimann during the intervening years, and found that even Minkowski’s new notion of space assumed too much. In short, Reimann demonstrates that a generalized n-dimensional manifold, if not assumed to be Euclidean in nature, can contain up to $(n)(n+1)/2$ terms in the general expression for distance. For instance, the $(2)(3)/2 = 3$ terms for distance on a generalized two-dimensional surface are contained in Gauss’ formula for a generalized, continuous linear element, which can be written in our terms as $ds = \sqrt{E\,dx^2 + 2F\,dx\,dy + G\,dy^2}$. Distance on a Euclidean plane surface can always be expressed, with the proper choice of coordinates, by $E = G = 1$ and $F = 0$, but the surface must necessarily be curved if it is not possible to choose coordinates so that the middle term drops out while the coefficients of the dx^2 and dy^2 terms become 1.

Extending this pattern to manifolds of higher dimension, it is clear that a three-dimensional manifold has $(3)(4)/2 = 6$ possible terms, of which the three non-squared differential terms ($dxdy$, $dxdz$, and $dydz$) are equal to zero for Euclidean space. Similarly, a four-dimensional manifold has $(4)(5)/2 = 10$ possible terms, of which six are zero in the Euclidean case. The general expression for distance in a four-dimensional manifold thus turns out to be

$$ds = \sqrt{E\,dx^2 + F\,dxdy + G\,dxdz + H\,dxdt + J\,dy^2 + K\,dydz + L\,dydt + M\,dz^2 + N\,dzdt + P\,dt^2},$$

of which the coefficients can be either constants or functions of the x, y, z, t. The notation can be simplified by putting the coefficients of each term in a 4×4 matrix with column indices σ and row indices τ:

$$g_{\sigma\tau} = \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & g_{34} \\ g_{41} & g_{42} & g_{43} & g_{44} \end{pmatrix},$$

which in Minkowski's version of spacetime is nothing but

$$g_{\sigma\tau} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Calling dx, dy, dz, dt instead dx_1, dx_2, dx_3, dx_4, we arrive at Einstein's notation for distance in a generalized four-dimensional manifold (119):

$$ds^2 = \sum_{\tau\sigma} g_{\sigma\tau}\, dx_\sigma\, dx_\tau.$$

Reimann also demonstrates (borrowing from Gauss) that "the intrinsic measure-relations of a twofold extent in which the line-element may be expressed as the square root of a quadric differential, which is the case with surfaces, are characterized by the total curvature" (7). This obtusely phrased claim essentially states that the curvature of a manifold is entirely determined by the distance formula appropriate to that manifold. The curvature of a particular species of non-Euclidean spacetime is then determined simply by the elements in the distance matrix above, turning the 0 and ± 1 terms into other constant coefficients or functions of position, that is, functions of x, y, z, and t. We do not have freedom to substitute freely, though, because distance should remain invariant regardless of the direction in which we choose to measure; thus, for example, $dxdz = dzdx$, and generally, the terms situated symmetrically about the diagonal from the upper left to the lower right must be equal. But the end result is astounding: All of nature is subsumed under geometry. Any acceleration comes to indicate a curve in space, and Einstein is able to demonstrate that a body seen to accelerate is in truth merely traversing a new shortest line, or "geodesic," in that particular region of necessarily-curved space. A fruitful treatment of these consequences is given in Einstein's *The Foundation of the General Theory of Relativity*.

It remains then only to confront Kant. If we grant that the *a priori* consistency of mathematics provides definitive generalizations about the nature of the Universe beyond our sense experience—that is, if we grant that mathematical results in cases like this achieve results more substantiated than simple arguments by analogy—then we can justifiably claim that Minkowski's construct, and Einstein's further development of it, constitute more than a mere mathematical model. The construct must necessarily correspond at least formally to the nature of the space we inhabit. To space is thus attributed specific *properties*, and we can no longer consider it as an empty stage upon which the bodies of the Universe interact. Space instead provides the conditions under which they are forced to interact. Reimann is able to draw from this a far-reaching cosmological consequence: "if we assume independence of bodies from position, and therefore ascribe to space constant curvature, it must necessarily be finite provided this curvature has ever so small a positive value. If we prolong all the geodesics starting in a given surface-element, we should obtain an unbounded surface of constant curvature, *i.e.*, a surface which in a *flat* manifoldness of three dimensions would take the form of a sphere, and consequently be finite" (10). In other words, if the terms in the $g_{\sigma\tau}$ matrix corresponding to this Universe indicate constant positive overall curvature, even with minor local fluctuations, its shape can be determined solely by mathematical operations on paper. That shape would be unbounded—Lucretius could never reach an edge from which to heave his spear—but finite in extent.

The problem arises in the imagination. In our flat three-dimensional imagination, envisioning a spherical surface presents no problem. But at the demand to imagine an unbounded, finite space—space which curves back on itself—the imagination blinks uncomprehendingly and shrugs. At the demand to imagine an analogous four-dimensional manifold, the imagination just laughs and walks away. This reaction confirms Kant's suspicion that human consciousness is resigned to eternally haunt its three-dimensional Euclidean world, at the same time that it undermines his claim that we will never perceive beyond the rigid confines of the pure forms of the sensible intuition. Mathematics describes for us a world we cannot imagine, but with mathematics as a lens by which we focus further inquiry, the task then lies in reordering and reorienting the imagination to encompass this new world.