Flat Space and Curved Space

Randy S

Abstract The familiar euclidean geometry of three-dimensional space is an example of **flat space**. This article introduces the concept of curved space, which is an easy generalization after we change the way we think about geometry in flat space.

Contents

1 Introduction

Understanding curved space is easier after we change the way we think about geometry in ordinary flat space:

> Think of geometry as something that assigns a length $L(P)$ to every path P.

We can recover everything else about the geometry from the function $L(P)$. We don't need to know in advance whether a path is straight or curved. The function $L(P)$ is defined for all smooth paths, and it tells us which paths are straight: Among all the paths from a to b, the one with the minimum length is the one we call straight.

Sections 2 and 3 explain how the function $L(P)$ can be described concisely. After we understand how to do this, the generalization from flat space to curved space is easy (section 11).

2 Three-dimensional flat space

A **coordinate system** is a way of labeling the points of space: each point of three-dimensional space is labeled by a triple of real numbers, (x, y, z), varying smoothly from one point to the next.[1] Given a coordinate system, we can specify any path using three functions

$$x(\lambda),\ y(\lambda),\ z(\lambda) \tag{1}$$

whose derivatives are not all zero for any value of the parameter λ. Different values of λ specify different points along the path, and the functions (1) give the coordinates of each of those points. The path is called **smooth** if the functions (1) have well-defined derivatives. A smooth path does not have any sharp kinks or discontinuities.

To endow three-dimensional space with geometry, we can specify a **line element**. The line element assigns a length $L(P)$ to every finite path P. For flat space, the line element defines a function $s(\lambda)$ by

$$\boxed{\dot{s}^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2.} \tag{2}$$

An overhead dot denotes a derivative with respect to λ:

$$\dot{x} \equiv \frac{dx}{d\lambda}. \tag{3}$$

Then the quantity

$$\big|s(\lambda_2) - s(\lambda_1)\big| \tag{4}$$

defines the length of the given path from λ_1 to λ_2. The next section explains why this works.

[1] We don't need to use "axes" to specify a coordinate system. Axes are for drawing graphs, not for formulating the laws of geometry or physics.

3 Why it works

To see that equation (2) defines the familiar geometry of flat space, recall the definition of the derivative:

$$\dot{x} \equiv \lim_{\Delta\lambda \to 0} \frac{x(\lambda + \Delta\lambda) - x(\lambda)}{\Delta\lambda}.$$

Intuitively, instead of taking the limit $\Delta\lambda \to 0$, we can think of $\Delta\lambda$ as a tiny but nonzero real number. Then the derivative is approximated by a finite difference:

$$\dot{x} \approx \frac{\Delta x}{\Delta\lambda}$$

with

$$\Delta x \equiv x(\lambda + \Delta\lambda) - x(\lambda).$$

Use this in (2) to get

$$(\Delta s)^2 \approx (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2. \tag{5}$$

We recognize this as the Pythagorean theorem for the given tiny segment of the path. We can subdivide the path into tiny segments, and then we can calculate the length of the whole path by summing the lengths of all of these tiny segments, using (5) for each one. The approach described in the previous section makes this idea precise, using derivatives to capture the limit $\Delta\lambda \to 0$ so that the approximation becomes exact.

The line element (2) is often written like this:[2]

$$\boxed{ds^2 = dx^2 + dy^2 + dz^2.} \tag{6}$$

This abbreviation can be motivated by thinking of it as the "numerator" of (2), after canceling the factors of $d\lambda^2$ in the "denominator" as explained above.

[2] The prefix "d" stands for "differential." Parentheses are implied, so that ds^2 is an abbreviation for $(ds)^2$.

4 Generalizations

This can all be generalized in several ways:

- It can be generalized to D-dimensional space, for any integer $D \geq 1$. This generalization is easy: just use D coordinates instead of 3 coordinates. This generalization will be introduced in section 9.

- It can be generalized to curved space. This generalization will be introduced in section 11.

- It can be generalized to spacetime. Then some paths have a length and some other paths have a duration. The line element specifies both, and it tells us which paths have which property. This generalization is introduced in article 48968.

- We can combine all of these generalizations to describe N-dimensional curved spacetime. This is introduced in article 48968.

But before we move on to those generalizations, let's use the special case that was introduced in section 2 to illustrate some important principles:

- The length of a path is independent of how the path is parameterized by λ. This is explained in section 5.

- The same space (flat three-dimensional space in this case) can be described using different coordinate systems, and the line element looks different in different coordinate systems. This is explained in sections 6-8.

5 Reparameterization invariance

The same path can be parameterized in different ways. As an example, consider the path

$$\begin{aligned} x(\lambda) &= \lambda \\ y(\lambda) &= 3\lambda \\ z(\lambda) &= \lambda^2 \end{aligned} \tag{7}$$

and the path

$$\begin{aligned} x(\lambda) &= \lambda + \lambda^3 \\ y(\lambda) &= 3(\lambda + \lambda^3) \\ z(\lambda) &= (\lambda + \lambda^3)^2. \end{aligned} \tag{8}$$

These are the *same path*, because we can convert the first one to the second one just by replacing λ with $\lambda + \lambda^3$, which is a monotonically increasing function of λ.

The fact that they both describe the same path should be obvious. If it doesn't seem obvious yet, remember that when we're specifying a path, the parameter λ is really just playing the role of an index (which we would often write as a subscript), even though it's continuous. When we replace λ with $\lambda + \lambda^3$, we're just using a different way of indexing the same set of points in space.

Since the same path can be parameterized in different ways, we might wonder if the length defined by equation (2) depends on how the path is parameterized. It doesn't. Intuitively, this is because the factors of $(\Delta\lambda)^2$ cancel: see section 3. The length defined by (2) only depends on the path itself, not on on how the path is parameterized.

We do need to be careful when specifying the endpoints of the path in equation (4). If we change how the path is parameterized, then we change which values of the parameter correspond to the path's endpoints. The statement that the length of the path is independent of the parameterization assumes that the path itself – including its endpoints – is fixed.

6 Coordinate invariance

A coordinate system is a way of assigning a unique label to each point of space. The length of a path does not depend on which coordinate system we use, but the equation for the length of a path may *look* different in different coordinate systems.

For an example, consider two-dimensional space. Then equation (2) reduces to

$$\dot{s}^2 = \dot{x}^2 + \dot{y}^2. \tag{9}$$

Define a new coordinate system (X, Y) by

$$X \equiv x/(1+y^2) \qquad\qquad Y \equiv y. \tag{10}$$

We can re-arrange this to get

$$x = (1+Y^2)X \qquad\qquad y = Y. \tag{11}$$

The chain rule (which should be familiar from calculus) implies

$$\dot{x} = (1+Y^2)\,\dot{X} + 2XY\,\dot{Y} \qquad\qquad \dot{y} = \dot{Y}.$$

Substitute these into (9) to get

$$\dot{s}^2 = (1+Y^2)^2\,\dot{X}^2 + 4XY(1+Y^2)\,\dot{X}\,\dot{Y} + \big(1 + (2XY)^2\big)\,\dot{Y}^2. \tag{12}$$

Equation (12) still defines the geometry of flat space. It looks different than (9) only because it's expressed in a different coordinate system. Starting with equation (12), we can recover (9) by using (10) to write (X, Y) in terms of (x, y).

Consider the path defined by $\big(x(\lambda), y(\lambda)\big) = (2\lambda, \lambda)$ for $0 \le \lambda \le 1$. According to equation (9), the length of this path is $\sqrt{5}$. According to equation (10), the same path can be described in the other coordinate system as $\big(X(\lambda), Y(\lambda)\big) = \big(2\lambda/(1+\lambda^2), \lambda\big)$ for $0 \le \lambda \le 1$. According to equation (12), the length of this path is $\sqrt{5}$. The length of a given path is the same no matter what coordinate system we use.

7 Coordinates don't always cover the whole space

A single coordinate system doesn't need to cover the whole space. We can use different (overlapping) coordinate systems to cover different parts of the space. This section shows an example of a coordinate system that only covers part of the space.

Start with two-dimensional flat space in the usual coordinate system:

$$\dot{s}^2 = \dot{x}^2 + \dot{y}^2. \tag{13}$$

For $x > 0$, define new coordinates X, Y by

$$x = e^X \qquad\qquad y = Y. \tag{14}$$

This new coordinate system only covers the half-space $x > 0$, because as X ranges over the whole real line $(-\infty, \infty)$, the value of $x = e^X$ ranges over the half-line $(0, \infty)$. Substitute (14) into (13) to get

$$\dot{s}^2 = e^{2X}\,\dot{X}^2 + \dot{Y}^2. \tag{15}$$

Or, in differential notation,

$$ds^2 = e^{2X}\,dX^2 + dY^2.$$

This is the line element for the half-space $x > 0$ in the new coordinate system. The new coordinates are unrestricted ($-\infty < X < \infty$ and $-\infty < Y < \infty$), but they only cover half of the space.

8 Example of a useful coordinate transformation

In the example shown in section 6, the relationship between (x, y) and (X, Y) was one-to-one. This ensures that, in either coordinate system, each point is labeled by a unique pair of numbers, and each pair of numbers labels a unique point.

Sometimes we may want to use a "coordinate system" in which the relationship between points and number-pairs is not one-to-one everywhere. If it is one-to-one within some restricted region of space, then this is still a legimate coordinate system within that region. An important example in two-dimensional space is the system of **polar coordinates** defined by the relationships

$$
\begin{aligned}
x &= X \cos Y \\
y &= X \sin Y.
\end{aligned}
$$

Writing r and ϕ instead of X and Y makes this look more familiar:

$$
\begin{aligned}
x &= r \cos \phi \\
y &= r \sin \phi.
\end{aligned}
$$

If we consider only the region of space covered by $r > 0$ and $0 < \phi < 2\pi$, then this is a one-to-one relationship, so we can use (r, ϕ) as a new coordinate system within that region. Use the identities

$$
\begin{aligned}
\dot{x} &= \dot{r} \cos \phi - \dot{\phi}\, r \sin \phi \\
\dot{y} &= \dot{r} \sin \phi + \dot{\phi}\, r \cos \phi
\end{aligned}
$$

in equation (9) to get

$$\dot{s}^2 = \dot{r}^2 + r^2 \dot{\phi}^2. \tag{16}$$

Or, in diffential notation,

$$ds^2 = dr^2 + r^2\, d\phi^2.$$

This coordinate system is called **polar coordinates**. Any coordinate system in which the length equation has the original form (9), or (19) in D-dimensional space, is called **Cartesian coordinates**. Most coordinate systems, like the one in section 6, do not have special names.

9 D-dimensional flat space

The generalization to D-dimensional space is easy. Each point of D-dimensional space can be labeled by a D-tuple of real numbers, $(x_1, x_2, ..., x_D)$, each of which varies smoothly from one point to the next.

Given a coordinate system, we can specify an arbitrary path using D functions

$$x_1(\lambda),\ x_2(\lambda),\ ...,\ x_D(\lambda) \tag{17}$$

whose derivatives are not all zero for any λ. As before, use an overhead dot to denote a derivative with respect to λ:

$$\dot{x}_n \equiv \frac{dx_n}{d\lambda}. \tag{18}$$

To define the geometry of flat D-dimensional space, we can use the line element

$$\boxed{\begin{array}{ll} \dot{s}^{\,2} = \dot{x}_1^2 + \dot{x}_2^2 + \cdots + \dot{x}_D^2 & \text{(derivative notation)} \\ ds^{\,2} = dx_1^2 + dx_2^2 + \cdots + dx_D^2 & \text{(differential notation)} \end{array}} \tag{19}$$

This generalizes equations (2) and (6). Just like (4), the quantity

$$\big|s(\lambda_2) - s(\lambda_1)\big| \tag{20}$$

defines the length of the part of the path that goes from λ_1 to λ_2. In this way, equation (19) implicitly specifies the geometry of D-dimensional **flat space**, by specifying the length of every smooth path.

10 General coordinate transformations

To describe a general coordinate transformation, let $(x_1, x_2, ..., x_D)$ denote the old coordinates of a point, and let $(X_1, X_2, ..., X_D)$ denote the new coordinates of the same point. We can write the old coordinates as functions of the new ones:

$$x_n(X_1, X_2, ..., X_D),$$

with one such function for each value of the index $n \in \{1, 2, ..., D\}$. We can abbreviate this function as $x_n(X)$. The chain rule gives the identity

$$\dot{x}_n = \sum_j \frac{\partial x_n}{\partial X_j} \dot{X}_j.$$

As before, an overhead dot means an ordinary derivative with respect to λ. Substitute this into (19) to get

$$\dot{s}^2 = \sum_{j,k} g_{jk} \, \dot{X}_j \dot{X}_k \tag{21}$$

with

$$g_{jk} = \sum_n \frac{\partial x_n}{\partial X_j} \frac{\partial x_n}{\partial X_k}. \tag{22}$$

Equation (21) can also be written in differential notation like this:

$$ds^2 = \sum_{j,k} g_{jk} \, dX_j \, dX_k.$$

The coefficients are functions of the coordinates X, but they don't depend on the derivatives $\dot{X}$. In other words, the coefficients g_{jk} may be different at different points in space, but they don't depend on which path we're considering.

Equations (21)-(22) imply the existence of a coordinate system in which the length equation has the simple form (19). As long as the coefficients have the form (22) with one-to-one functions $x_n(X)$, the geometry defined by (21) is flat.

11 Curved space

Inspired by (21), let's write the length equation as

$$\boxed{\dot{s}^2 = \sum_{j,k} g_{jk}(x)\,\dot{x}_j\dot{x}_k \qquad \text{or} \qquad ds^2 = \sum_{j,k} g_{jk}(x)\,dx_j\,dx_k} \tag{23}$$

from now on. The coefficients $g_{jk}(x)$ are the components of the **metric field**, also called the **metric tensor** or just the **metric**.

Depending on the coefficients $g_{jk}(x)$, equation (23) may or may not define the geometry of flat space. It might be the geometry of flat space in a coordinate system that makes it look more complicated (like equation (12)), or it might be something truly different – something that cannot be reduced to equation (19) no matter what coordinate system we use. The idea of **curved space** is that we can still use equation (23) to define the length of every path as long as the right-hand side is positive (not zero) whenever at least one of the $\dot{x}_j$ is non-zero. In other words, the right-hand side should be positive for every path. If the coefficients $g_{jk}(x)$ satisfy this condition, then equation (23) defines a self-consistent geometry.[3] The geometry that it defines may or may not be flat.[4]

This distinction between flat space and curved space refers to the *intrinsic* geometry of the space, which assigns a length to every path. This concept of curvature doesn't refer to any embedding of the space into a higher-dimensional flat space. General theorems about the existence of such embeddings have been proven,[5] but defining a manifold's *intrinsic* geometry does not require any such embedding.

[3] A smooth manifold equipped with such a metric is called a **Riemannian manifold**. Article 48968 considers metrics that are not positive-definite, in which case the manifold is called **pseudo-Riemannian**.

[4] It may be (exactly) flat in some places and not in others. Don't confuse this with the **local flatness theorem** (article 48968), which says that the geometry always looks *approximately* flat in a sufficiently small neighborhood of any given point.

[5] The **Nash embedding theorem** says that any Riemannian metric can be induced by embedding the manifold into a flat space of sufficiently many dimensions (Lee (1997), chapter 5, pages 66-67).

12 Most geometries are not flat

In the original special case (19), the components of the metric tensor are

$$g_{jk} = \delta_{jk} \equiv \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

To see that most geometries (23) are not flat, consider a metric tensor of the form

$$g_{jk}(x) = \delta_{jk} + \epsilon\, h_{jk}(x) + O(\epsilon^2) \tag{25}$$

with $|\epsilon| \ll 1$. The geometry might still be exactly flat even if $g_{jk} \neq \delta_{jk}$. However, if it is flat, then we know from section 10 that the coefficients g_{jk} can be written in the form[6]

$$g_{jk}(x) = \sum_n \frac{\partial X_n}{\partial x_j}\frac{\partial X_n}{\partial x_k} \tag{26}$$

for some set of functions $X_n(x)$. Since we're considering only slight deviations from the obviously-flat case, we can consider a coordinate transformation that only changes the coordinates slightly. For such a coordinate transformation, we can write the functions $X_n(x)$ as

$$X_n(x) = x_n + \epsilon\, z_n(x)$$

with $|\epsilon| \ll 1$. This gives

$$\partial_j X_n \equiv \frac{\partial X_n}{\partial x_j} = \delta_{jn} + \epsilon\, \partial_j z_n,$$

and using this in (26) gives

$$g_{jk} = \delta_{jk} + \epsilon\, h_{jk} + O(\epsilon^2)$$

[6] The letters x and X are switched, but otherwise this is the same as equation (22).

with

$$h_{jk} = \partial_j z_k + \partial_k z_j \tag{27}$$

Altogether, this shows that if the metric is flat, then the h in (25) can be written in the form (27). If the h in (25) cannot be written in this form, then the metric is not flat.

Given some set of functions $h_{jk}(x)$, how do we know if they can be written in the form (27)? Here's an easy test. For arbitrary functions z_n, the quantities (27) satisfy

$$\sum_{a,b,j,k} A_{aj} B_{bk}\, \partial_a \partial_b h_{jk} = 0 \tag{28}$$

whenever A and B are both antisymmetric, which means

$$A_{aj} = -A_{ja} \qquad\qquad B_{bk} = -B_{kb}.$$

The identity (28) is a straightforward consequence of the elementary identities $\partial_a \partial_j = \partial_j \partial_a$ and $\partial_b \partial_k = \partial_k \partial_b$. If we can find any antisymmetric A and B for which h_{jk} fails to satisfy (28), then the geometry with $\epsilon \neq 0$ is not flat. The example in the next section illustrates this.

By the way, the antisymmetrized combinations appearing in the condition (28) are the components of the small-ϵ approximation to the **curvature tensor**. A geometry is flat if and only if its curvature tensor is zero.

13 Example of a non-flat space

Use the abbreviation

$$r^2 \equiv \sum_k x_k^2$$

and consider the example

$$g_{jk}(x) = (1 + \epsilon r^2)\delta_{jk}. \tag{29}$$

In other words,

$$h_{jk} = r^2 \delta_{jk}.$$

This gives

$$\partial_a \partial_b h_{jk} = 2\delta_{ab}\delta_{jk}.$$

Now, choose any two vectors v and w and take

$$A_{aj} = B_{aj} = v_a w_j - v_j w_a.$$

Then

$$\sum_{a,b,j,k} A_{aj} B_{bk}\, \partial_a \partial_b h_{jk} = 4(v \cdot v)(w \cdot w) - 4(v \cdot w)^2, \tag{30}$$

with $v \cdot w \equiv \sum_k v_k w_k$ as usual. We can choose v and w so that $v \cdot w = 0$ and $v \cdot v = w \cdot w = 1$, in which case the quantity (30) is not zero. This proves that the geometry defined by (29) is not flat if $\epsilon \neq 0$.

14 A metric tensor is not just a set of components

We can think of $g_{ab}(x)$ as the components of the metric field, but a metric field is not just a collection of components. A metric field defines a *map* from one set to another, specifically from the set of world-lines smoothly parameterized by λ to the set of functions $s(\lambda)$ of λ. Both of these sets have coordinate-independent definitions, so the metric field is also a coordinate-independent entity: it is a map from one coordinate-independent set to another. The components $g_{ab}(x)$ depend on which coordinate system we use, but the map does not. In fact, the statement that the map does not depend on which coordinate system we use tells us how to transform the components $g_{ab}(x)$ when switching from one coordinate system to another. This was illustrated in sections 6-8 and 10.

Article 09894 explains how to define a metric field (and other tensor fields) without using coordinates at all.

15 The summation convention

The notation that was used in this article, using subscripts to distinguish the different coordinates from each other, is sufficient for many purposes. For other purposes, a different notation is more convenient. Instead of writing the coordinates index as a subscript, we can write it as a superscript, like this:

$$(x^1,\, x^2,\, ...,\, x^D).$$

We need to remember that when a coordinate is written as x^a, the superscript is an index, not an exponent.[7] Partial derivatives with respect to the coordinates are then abbreviated like this:

$$\partial_a \equiv \frac{\partial}{\partial x^a}.$$

A subscript is used for ∂_a because x^a is in the denominator, loosely speaking.

In the **(Einstein) summation convention**, a sum is implied whenever the same index appears as both a superscript and subscript in the same term. Example:

$$df = dx^a \partial_a f$$

with an implied sum over a. This convention is convenient because the combination $dx^a \partial_a$ (with as sum over a) is invariant under coordinate transformations: under a coordinate transformation $x \to X$, the factors of $\partial X/\partial x$ and $\partial x/\partial X$ cancel each other. Writing an index as a subscript or superscript conveys which of these two factors is involved in a coordinate transformation, so the positions of the indices tell us when such cancellations will occur. Another example:

$$ds^2 = g_{ab}(x) dx^a\, dx^b, \tag{31}$$

with implied sums over a, b. The fact that each index on the right-hand side occurs both as a subscript and as a superscript tells us that the factors of $\partial X/\partial x$ and $\partial x/\partial X$ cancel each other in a coordinate transformation (section 10).

[7] Superscripts are *also* used for exponents, so we need to pay attention to the context.

16 References

Lee, 1997. *Riemannian Manifolds: An Introduction to Curvature.* Springer

17 References in this series

Article **09894** (https://cphysics.org/article/09894):
"Tensor Fields on Smooth Manifolds" (version 2023-11-12)

Article **48968** (https://cphysics.org/article/48968):
"The Geometry of Spacetime" (version 2022-10-23)

Article **93169** (https://cphysics.org/article/93169):
"Derivatives and Differentials" (version 2023-05-14)