HOW TO THINK ABOUT POINTS AND VECTORS WITHOUT COORDINATES

Math 225

Points

Look around. What do you see? You see objects: a chair here, a table there, a book on the table. These objects occupy locations, or points. A physical object actually occupies infinitely many points, but typically you can identify particular points associated with an object: the corners of the table, the center of mass of the chair, the exact location in space of the center of the period at the end of the first sentence on page 43 of the book as it sits on the table. The last example can be talked about precisely even though it would be hard to pinpoint the location of the period—as soon as you pick up the book and turn to page 43 you move the period to a new location. We can give names to points that we want to talk about: P for the location of the period (or at least where it was before you picked up the book), Q, R, S, and T for the corners of the table, C for the center of mass of the chair. Given these points we can talk about things they determine: the line $\overleftrightarrow{PQ}$, the line segment $\overline{RT}$ (a diagonal of the table), the triangle $\triangle PTC$, or the plane containing this triangle.

Even though General Relativity tells us otherwise, let's pretend that we live in a three-dimensional Euclidean space, which we will denote by E^3 (the exponent stands for the dimension of the space, not an algebraic operation). The points in space are elements of the set E^3: $P, Q, R, S, T, C \in E^3$. The line segments, lines, triangles, planes and other objects they determine are subsets of E^3.

If we restrict our attention to a particular plane, say the table top or the chalkboard, we can refer to it as E^2, a two-dimensional Euclidean plane. Given two points $A, B \in E^2$ you can talk about the circle centered at A and passing through B.

Certain numbers are associated with these objects: the length of $\overline{RT}$ (which is also the distance between R and T), the area of $\triangle PTC$, the circumference of the circle centered at A that passes through B, the angle of inclination above the table that P makes when viewed from the corner Q of the table, measured in degrees or radians.

None of this should seem strange. This is how you thought about geometry in high school.

Now think for a moment about what you *don't* see. When you look at the center A of the circle in E^2, you don't see an ordered pair of numbers (like $(3, 2)$). When you think about the center of mass C of the chair, you don't think about an ordered triple of numbers (like $(5, -1, 2)$). In other words, you don't see coordinates. Similarly, you don't see equations for the circle or line.

This is not to say that coordinates aren't useful. They can be very useful. If you really want to talk about P carefully, it is helpful to take a coordinate system with origin at one of the corners of the table, two axes along edges of the table, and the third axis perpendicular to the table. Once the coordinate system is established, you can measure carefully and find the coordinates of P relative to this coordinate system. If someone else has done this, you may want to share information. But you may be using different coordinate systems! Some things you share will agree, such as distances, areas, angles. Other things will disagree, namely the coordinates of points and equations of lines. The things you agree on are purely geometric and independent of any coordinate system you might use. The things you disagree on are artifacts of the coordinate systems you are using. If you need to communicate things that depend on a coordinate system, you either need to agree on a particular coordinate system to use, or find formulas that will allow you to switch between the different coordinate systems.

A choice of coordinate system usually depends on the problem being studied. If you are thinking about a circle in E^2, you may want to choose the origin to be the center of the circle. Why? Because the equation for the circle is simplest, namely $x^2 + y^2 = 25$, if the radius is 5, as opposed to $(x-3)^2 + (y-2)^2 = 25$ if the origin is somewhere else (namely, the point with coordinates (3,2)). Suppose you carry the chair up to the roof of Goodrich Hall and throw it over the edge, making it tumble as it falls. In this case it's handy to use a *moving* coordinate system in which the origin is the center of mass of the chair and the coordinate axes rotate with the chair as it spins.

There are other kinds of coordinate systems. The equation for a circle of radius 5 is very simple if you use polar coordinates with the origin at the center of the circle, namely $r = 5$. Similarly, in three dimensions the use of cylindrical or spherical coordinates often helps in certain situations. In this class we will have the opportunity to use coordinates in which the coordinate "lines" are parabolas, ellipses, or hyperbolas!

We are used to using language like "the point P $(3, 5, -1)$." Of course what we really mean is that the coordinates of P are $(3, 5, -1)$ relative to some coordinate system we have in mind. Similarly, we talk about $\mathbb{R}^3$ as being space, when in actuality, $\mathbb{R}^3$ is the collection of all coordinates: $\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$. *The identification of $\mathbb{R}^3$ with three-dimensional space is the choice of a coordinate system.* These comments are obvious when pointed out this explicitly, but are easy to forget when involved in the details of an example or problem.

Some objects are in motion: an airplane, a flying insect, your friend walking across the room. A moving particle in the plane can be described as a function of the form $\gamma\colon \mathbb{R} \to E^2$. If $t \in \mathbb{R}$ is a particular time, then $\gamma(t) \in E^2$ is the location of the particle at that time. If we are using a coordinate system, the motion can be expressed in terms of the coordinates. At time t, denote the coordinates of the particle by $(x(t), y(t))$. Then $x, y\colon \mathbb{R} \to \mathbb{R}$ are the coordinate functions. For example, $(x(t), y(t)) = (3\cos t, 2\sin t)$ is the coordinate description of a particle moving counter-clockwise around an ellipse. When we are being casual we might write $\gamma(t) = (x(t), y(t)) = (3\cos t, 2\sin t)$, but notice that this is the same error we make when we talk about "the point P $(3, 5, -1)$," namely, it is an identification of the location, $\gamma(t)$, of the particle with the coordinates, $(x(t), y(t))$, of its location. Much of the time there is no harm in this, unless the identification becomes so ingrained that we have trouble making the distinction.

Vectors

When you look around you, points are easy to see. A point is just an idealized notion of position. You start seeing vectors when you think about how to get from one point to another. Let P and Q be two points in E^2, and consider the segment $\overline{PQ}$. The idea of moving along this segment from P to Q gives the segment a direction, or orientation, and this is one way to think about the vector $\overrightarrow{PQ}$. We will call this vector the *displacement vector* from P to Q. This same vector can be used to displace any other point R in the plane. If P, Q, and R are not colinear, then there is a fourth point S so that the quadrilateral $PQSR$ is a parallelogram (draw a picture!). Then $\overrightarrow{RS} = \overrightarrow{PQ}$, that is, the displacement that takes P to Q also takes R to S. (Think about how to do this if P, Q, and R lie on the same line.) Thus the vector $\mathbf{v} = \overrightarrow{RS} = \overrightarrow{PQ}$ determines a *rigid motion* of the entire plane. Each point is displaced along a line parallel to $\overleftrightarrow{PQ}$ in the same direction and with the same distance. Note that displacement vectors depend on the Euclidean geometry of E^2, but not on coordinates—you don't have to think about ordered pairs of numbers to visualize an arrow from one point to another.

You may be used to locating points by using position vectors. However the use of position vectors requires some particular point O to be the origin. The point P is identified with the vector $\overrightarrow{OP}$. A different choice of origin would identify P with a different vector. If you don't have an origin, you can still think about points, but you can't think about them as vectors. *The identification of points and vectors is origin-dependent.* It's better to think about points and vectors separately, and so I tend to avoid position vectors. Velocity vectors, on the other hand, are origin-independent, as we will see. The upshot of this is that single points do not determine vectors. Rather, *pairs* of points determine vectors.

Note that everything said in this section applies to displacement vectors in E^3 as well.

The Geometric Algebra of Points and Vectors

The collection of displacement vectors (whether we are in E^2 or E^3) forms a vector space. Vector addition satisfies the parallelogram law, namely, $\overrightarrow{PQ}+\overrightarrow{QS} = \overrightarrow{PR}+\overrightarrow{RS} = \overrightarrow{PS}$ (draw the picture). Scalar multiplication also satisfies some obvious properties. For example, if M is the midpoint of $\overline{PQ}$, then $\overrightarrow{PM} = \frac{1}{2}\overrightarrow{PQ}$. Also, $\overrightarrow{QP} = -\overrightarrow{PQ}$.

The difference between E^2 and E^3 is the dimension of the space of displacement vectors. The dimension for E^2 is 2, for E^3 it is 3. This should be fairly intuitive. If P, Q, and R are three non-colinear points in E^2, then the vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ form a basis. Similarly, if P, Q, R and S are four non-coplanar points in E^3, then the vectors $\overrightarrow{PQ}$, $\overrightarrow{PR}$, and $\overrightarrow{PS}$ form a basis.

It turns out to be very useful if we extend the notions of addition and subtraction of vectors to certain combinations of points and vectors. In particular, it's nice to think of the process of applying the displacement vector $\overrightarrow{PQ}$ to the point P (resulting in the point Q) as *adding* $\overrightarrow{PQ}$ to P.

Definition. The sum of P and $\overrightarrow{PQ}$ (written $P + \overrightarrow{PQ}$) is the point Q. More generally, if P is any point and $\mathbf{v}$ is any vector, let Q be the point such that $\mathbf{v} = \overrightarrow{PQ}$, and define $P + \mathbf{v}$ to be Q. (You need to draw the picture to make sense of this.)

Suppose P is a point and $\mathbf{v}$ and $\mathbf{w}$ are two vectors. With the definition of addition just given, it can be shown that

$$(P + \mathbf{v}) + \mathbf{w} = P + (\mathbf{v} + \mathbf{w}) = (P + \mathbf{w}) + \mathbf{v}.$$

These really mean different things! Draw a picture! In all three you are going from one corner of a parallelogram, namely P, to the opposite corner, but you are taking differernt routes. In two of the expressions you are following the two different paths along the edges of the parallelogram. In the remaining expression you are going along the diagonal. Which is which? Note that the sums in the first and third expressions are all between points and vectors. In the second expression, the first sum is between a point and a vector, whereas the second sum is between two vectors, that is, it's the addition in the vector space. Since these three expressions are all equal, we can drop the parentheses and simply write $P + \mathbf{v} + \mathbf{w}$.

If $P + \mathbf{v} = Q$, it seems reasonable to have $\mathbf{v} = Q - P$. This subtraction requires a definition also.

Definition. If P and Q are points, then $Q - P$ is defined to be the displacement vector $\overrightarrow{PQ}$.

If $P + \mathbf{v} = Q$, then $\mathbf{v}$ is $\overrightarrow{PQ}$, by the definition of addition. By the definition of subtraction, $\mathbf{v}$ is also $Q - P$. Note this says that in an equation of points, you can subtract a point from both sides, resulting in an equation of vectors.

It turns out that the familiar vector-algebraic operations (adding or subtracting the same quantity to both sides of an equation, or multiplying both sides of an equation by the same scalar) are legitimate, **provided**

(1) We never add two points, and
(2) We never multiply a point by a scalar.

The reason is that these operations don't make sense geometrically!

If all of this seems like just so much playing with notation, here are some things to consider. What we have defined is an algebraic way to talk about how points and displacement vectors interact, and it works without referring to coordinates or position vectors. In the next section we will see that if we do use coordinates or position vectors, this algebra is exactly the same as what you are used to. The gain, which is not easy to appreciate when you first see this, is conceptual. This point of view leads to a deeper understanding of two and three-dimensional space, and of the problems in these spaces that we want to understand. We will be able to talk about problems in a way that minimizes the mention of things that are incidental (coordinates of points and components of vectors relative to the standard basis for example), thereby focusing on the things that are central.

The Full Circle—Coordinates and Position Vectors

Okay, suppose that you *are* working a problem with coordinates, which you are also used to treating as position vectors. How are the operations you are used to related to those just defined?

Here is what you are used to, treating everything as vectors: if $\mathbf{p}$ and $\mathbf{q}$ are $(3, 2)$ and $(5, 1)$, respectively, then the vector from $\mathbf{p}$ to $\mathbf{q}$ is $\mathbf{q} - \mathbf{p} = (2, -1)$. Also, if you move from $\mathbf{p}$ to $\mathbf{r}$ using the vector $\mathbf{v} = (4, -3)$, you get $\mathbf{r} = \mathbf{p} + \mathbf{v} = (7, -1)$.

There are three observations to make here.

(1) This works if you are using Euclidean coordinates (the (x, y) or (x, y, z) coordinates you usually use in which the coordinate axes are perpendicular and the "unit" markings along the axes have the same length), but not polar coordinates.
(2) This works in *any* rectangular coordinate system, in which the axes are perpendicular, but the unit markings on different axes can be spaced differently.
(3) In fact, this works in any *linear* coordinate system (another name is *affine* coordinate system), in which the coordinate axes don't even need to be perpendicular, although they still need to be lines.(A linear, but non-rectangular coordinate system is the analog of having a basis for a vector space that isn't orthonormal.)

Since the coordinates and position vector of a point depend on which coordinate system is being used, I don't like to say that P *is* $(3, 2)$ or that $(3, 2)$ is a point. It's better to say that we are using a particular coordinate system, and in that system the coordinates of P are $(3, 2)$.

Since we are distinguishing between coordinates and vectors, it's good to have distinct notation for these concepts. Thus the position vector of P is $3\mathbf{i} + 2\mathbf{j}$, where $\mathbf{i}$ and $\mathbf{j}$ are the basis vectors for the coordinate system, *i.e.*, $\mathbf{i}$ is the vector from the origin to the point with coordinates $(1, 0)$, and $\mathbf{j}$ is the vector from the origin to the point with coordinates $(0, 1)$. Note that $\overrightarrow{PQ}$ is a vector, not a point, and so we write it as $2\mathbf{i} - \mathbf{j}$, not as $(2, -1)$. (Note that $\mathbf{i}$ and $\mathbf{j}$ need not be orthogonal or unit vectors! This can happen if the coordinate system is non-Euclidean.)

With this different point-of-view, the statement above becomes: if P and Q are points with coordinates $(3, 2)$ and $(5, 1)$, respectively, then the vector $\overrightarrow{PQ}$ is $2\mathbf{i} - \mathbf{j}$, that is, to get the *coefficients* of the vector from P to Q you subtract the *coordinates* of P from the *coordinates* of Q. If you apply the displacement vector $\mathbf{v} = 4\mathbf{i} - 3\mathbf{j}$ to the point P with coordinates $(3, 2)$, you get the point R with coordinates $(7, -1)$, that is, you add the *coefficients* of the vector to the *coordinates* of the point to get the *coordinates* of the new point.

These examples illustrate the reasons we called the new operations (defined in the previous section) addition and subtraction, instead of something else. The operations are closely related to addition and subtraction of vectors, and in fact, they are the same as addition and subtraction if you identify (I would say "confuse") points with their position vectors.

To be less wordy, we will write these operations as

$$Q - P = (5, 1) - (3, 2) = 2\mathbf{i} - \mathbf{j} \qquad \text{and} \qquad P + \mathbf{v} = (3, 2) + 4\mathbf{i} - 3\mathbf{j} = (7, -1) = R.$$

Technically speaking, this is an abuse of notation, since P, Q, and R really *aren't* $(3, 2)$, $(5, 1)$, and $(7, -1)$, rather the latter are the coordinates of the former. Nonetheless, this notation carefully distinguishes between points and vectors, which is the main point.

A Rigorous Approach to These Ideas

The previous sections are not rigorous mathematics. To be rigorous, the terms need to be precisely defined and the properties proved. On the other hand, the ideas presented are how mathematicians intuitively think about these things.

A Final Word

The title of this handout is a bit misleading. We aren't going to think entirely without coordinates. The point is to separate the geometry from the coordinates. The important facts that we are interested in studying are independent of any coordinate system. In some problems a coordinate system is very useful, particularly if detailed computations are needed. Believe it or not, we will be able to solve some problems completely without using coordinates! Most often, the solution without coordinates is easier than the one with coordinates! A better title might be "A coordinate-independent way to think about points and vectors."

The choice of a coordinate system is somewhat arbitrary. A "good" coordinate system for a particular problem will reflect some symmetry or other aspect of the problem. When a coordinate system is used to solve a problem, it's good to think about what parts of the solution depend on the coordinates and what parts don't. The important quantities will be the ones that don't depend on the coordinates.

Here's an idea of what I mean. Suppose you are thinking about a physics problem in which some particle mass is moving around in a plane subject to some force. The point of the problem might be to find the particle's velocity at some moment. The problem may specify a coordinate system, or you may be expected to choose one. The usual process is to resolve all vectors into their x and y components, and to make all computations using these components. If v_x and v_y denote the components of velocity, your answer might be $v_x = 3$ and $v_y = 4$. The numbers 3 and 4 by themselves aren't all that important, since they depend on the coordinate system. What *is* important is the velocity vector, which is the linear combination $\mathbf{v} = 3\mathbf{i} + 4\mathbf{j}$. This linear combination *does not* depend on the coordinate system. Here's what this means. Suppose another person solves the problem with a different coordinate system. This coordinate system will have different basis vectors, which we might denote by $\mathbf{I}$ and $\mathbf{J}$. The components v_x and v_y the second person gets might be $1+2\sqrt{3}$ and $2-\sqrt{3}$. If both people have done the problem correctly, they will find that $3\mathbf{i} + 4\mathbf{j} = (1 + 2\sqrt{3})\mathbf{I} + (2 - \sqrt{3})\mathbf{J}$. Thus they really do get the same answer. Their answers just *look* different. (The lesson to be learned from this is that when you get $v_x = 3$ and $v_y = 4$, you're not quite done. Your answer should be $3\mathbf{i}+4\mathbf{j}$.) A third person might come along and solve the problem without using coordinates at all (this is possible for some problems), getting some expression for $\mathbf{v}$. Assuming no errors have been made, if he expresses his $\mathbf{v}$ as a linear combination of $\mathbf{i}$ and $\mathbf{j}$, he will get $\mathbf{v} = 3\mathbf{i} + 4\mathbf{j}$.

Robert L. Foote
September 1998
August 2006