On the geometry of electromagnetism

Alain Bossavit, Électricité de France
1, Av. Gal de Gaulle, 92141 Clamart
Alain.Bossavit@edf.fr

(1): Affine space

INTRODUCTION

What is this series[*] of articles about? Examine the following figure, which displays Maxwell equations in three different formalisms. Never mind what these equations mean (this is summarized in the caption, but it's a secondary point). My immediate purpose is to call attention on how they *look*. In spite of describing the same physical phenomena, they are as different as three sentences with the same meaning can be in three different languages. The fact that one can discuss the same physics within widely different mathematical formalisms is what will concern us here. In particular, I wish to show that different *geometrical objects* can serve in describing electromagnetism: vector fields, differential forms, even quaternions, as in Maxwell's time, and so forth: "axial" vectors, "polar" vectors .... There is terrible confusion around the latter concepts, which I hope to dispel a little by showing how they relate. This will be, if not the exclusive subject, at the very least the red thread connecting these columns.

The linguistic metaphor should not be stretched too far, but it's apt to some extent. You can say one thing in Japanese and exactly the same thing in English, in most scientific contexts (leaving apart of course, poetry, allusions to political actuality, and jokes). You will not just substitute word for word, however, for both languages use different grammatical categories. Box a of Fig. 1 reproduces something Maxwell told us about the way the world behaves [Ma]. Boxes b and c show how this message translates in two contemporary languages: the "vector fields" formalism of most textbooks and Journals (box b) and the "four-dimensional differential forms" idiom (box c) of many Physics treatises (see, e.g., [Mi]). As one sees, the translation does not consist in a mere change of notation (passing from $\mathcal{H}$ to **H**, for instance), for even when the symbols look alike, they denote different kinds of entities—different "geometrical objects".

$$4\pi C = \mathrm{v}\cdot\nabla \mathrm{H}, \qquad C = J + \dot{D},$$
$$B = \mathrm{v}.\nabla U, \qquad E = -\dot{U} - \nabla\Psi,$$
$$B = \mu H, \qquad D = (4\pi)^{-1}\kappa E$$

(a)

$$-\partial_t \mathbf{D} + \mathrm{rot}\,\mathbf{H} = \mathbf{J}, \qquad \partial_t \mathbf{B} + \mathrm{rot}\,\mathbf{E} = 0,$$
$$\mathbf{B} = \mu\mathbf{H}, \qquad \mathbf{D} = \varepsilon\mathbf{E}$$

(b)

$$\mathrm{dF} = 0, \quad \mathrm{G} = *\mathrm{F}, \quad \mathrm{dG} = \mathrm{J}$$

(c)

Figure 1. Maxwell equations, with given currents: in the style of Maxwell's treatise (box a), in what may be the most widely accepted contemporary formalism (box b), and in modern differential geometric notation (box c). Maxwell used quaternions. (The v means "vector part" of a quaternionic product, and ∇ is the operator $\mathrm{i}d/dx + \mathrm{j}d/dy + \mathrm{k}d/dz$.) Notwithstanding, his formalism is not so far from today's received notation (box b), in which, apart from factors 4π, box a would read $\mathrm{rot}\,\mathbf{H} = \mathcal{C} = \mathbf{J} + \partial_t\mathbf{D}$, $\mathbf{B} = \mathrm{rot}\,\mathbf{A}$, $\mathbf{E} = -\partial_t\mathbf{A} - \nabla\psi$. Boxes b and c say exactly the same thing, but whereas **D**, **B**, etc., denote vector fields, F, G, and J are differential forms (J combines electric current density and electric charge in a single entity, F is "Faraday's tensor" and the star is the so-called "Hodge operator" in Minkowski's metric).

Maxwell's apparently bewildering notation begins to make sense when one remembers he had adopted *quaternions* as basic entities (at least, at the beginning; later, this changed; see e.g., [Cr] or [Sp] for historical accounts

[*] Published in **J. Japan Soc. Appl. Electromagn. & Mech., 6** (1998), pp. 17-28 (no 1), pp. 114-23 (no 2), pp. 233-40 (no 3), pp. 318-26 (no 4).

of the evolution towards the vector formalism of today). Similarly, we shall discover the kinship between (b) and (c) by examining the relations that exist between differential forms and vector fields. Let me hasten to say that quaternions will not be addressed here. Not that they should be confined to the dustbins of history, far from it. (They are quite useful in modern work on robotics, for instance.) But their relevance to electromagnetism was 19th-century illusion. Differential forms, on the other hand, are the right stuff—as I hope to show. But the way they are introduced in classics of differential geometry (which almost all discuss box c, if only in a rather thin chapter) cannot be recommended to Engineers. We shall adopt a different approach, leading to a formalism much closer to the familiar one of box b, by setting aside time and 3D-space, which the four-dimensional equations (c) do not distinguish.

In order so to stay very close to the familiar concepts, we shall have to introduce, and discuss with care, an appropriate *geometrical framework*, consisting of things such as *affine* three-dimensional *space*, the associated vector space, and geometrical objects living therein. (To "live" means that such entities may assume varying values as time goes on.)

That such a critical discussion be necessary at all is not so obvious. We often take for granted 3D space (good old Euclidean space, that is), as the natural framework in which to do physics, and though we shall not depart from this tradition here, I want to stress that this is a *modelling decision*, something that is to a large extent *up to us*, human beings, not something forced on us by the very structure of the Universe. The World *is*, and it certainly has order and structure. But order and structure in our *descriptions* of the world are something else, even if we try our best towards a close match, in the process of *model building*.

This activity—model building—is what distinguishes "pure" mathematics from "applied" ones. Pure mathematicians try to discover, analyse, and classify *all* logically possible abstract structures. People who apply mathematics, including physicists and engineers, use them to construct *specific* abstract structures, which reproduce some of the features of the real world, and thus can help in explaining or predicting the behavior of some definite segment of reality.

So mathematical entities by which we thus describe physics are not a priori frames of our thinking. They are our creation, moulded of course by the structures of the world out there, but still *abstract* things. Therefore, they are more or less adequate as tools with which to deal with the real world, which means one can—and one should—*criticize* the way they are applied, and question their adequacy. This process of critical reevaluation (constantly reinvigorated by new engineering practices, such as programming and computing) is the impetus that forces formalisms to evolve, even in well-understood compartments of physics, like classical electromagnetism, as witnessed by Fig. 1.

The purpose of these articles is thus to critically examine the geometrical concepts which compose the current formalism of electromagnetism. Hence a discussion at two levels: the formal one of mathematics, where one introduces abstractions (such as, for instance, three-dimensional affine space), and the practical one, where one passes judgment on their relevance to model building. This explains the alternation, in these columns, between descriptions of geometrical objects, and discussions of their physical significance.

And now, rather than go on philosophizing, let's do it.

1. AFFINE SPACE

Nothing is built without foundations, so we shall assume some preliminary knowledge: sets, functions and maps, elementary logic, and some familiarity with the basic structures: group, field, vector space .... Recall that all such structures are sets,[1] but not naked sets: structure is conferred on such sets by specific systems of relations and operations, which tell what can be done with and to the elements of the set.

1.1 Vector spaces

For instance, a *vector space*[2] on the reals is a set of objects called *vectors*, which one can

[1] One can (though this is not the only way) present mathematics in such a light that *all* mathematical objects are sets of some kind.

[2] Defined terms are set in *slanted* style, on first appearance. [Footnote's footnote, Sept. 2002: Fonts have changed, with respect to the original, in what you are

(1) add together (e.g., forming vector $v + w$ from vectors v and w) and (2) multiply by real numbers (e.g., forming vector λv from vector v and real number λ). No need to recall the properties required of these two operations, if the set is to qualify as a vector space. Just be aware that "vector" is a generic name, which may apply to other objects than the familiar two- or three-dimensional vectors of elementary geometry, provided the set of all objects thus considered obeys the vector-space axioms.

For instance, think of the electromagnetic (EM) field, at any instant, in one of these large experience halls in which high-voltage electrical hardware can be tested. At this stage of the discussion, we pretend not to know what the EM field "is", meaning that we are not yet committed to a specific mathematical object by which to *model* this empirical reality, the physical EM field (detectable by its effects on dust, on our hair, etc.). But we know that two EM fields can be superposed, adding up their effects, and that a given field can be scaled up by a factor 2, say, giving conceivable[3] EM fields. So it makes sense[4] to consider the set of all conceivable EM fields, in this hall, as a vector space. One can then envision the evolution of the experiment in the hall as a (continuous) sequence of values of a representative vector in this abstract space, in other words, as a trajectory. And there we are, with the beginning of a *geometrization* of the whole thing.

The main feature which distinguishes such vector spaces of fields (often called "functional spaces") from those of plane or spatial geometry is, of course, dimension. The *dimension* of a vector space is the maximal number of linearly independent vectors, if there is such a maximum (otherwise we have an infinite dimensional space, like the above space of EM fields). A *basis*, or *frame*, in a vector space V of finite dimension n is a family of n linearly independent vectors. Applied mathematicians seem to be especially fond of a particular space of dimension 3, denoted $I\!R^3$. This is the set of all triples of real numbers $\{x, y, z\}$. Such triples can be added or scaled up the obvious way. Reading such paper titles as "MHD in a subset of $I\!R^3$", or "Wave propagation in a stratified region of $I\!R^3$", one might believe that this particular vector space is the natural framework in which to do physics, which I think is silly, and is one of the received ideas I want to challenge here. (But one thing at a time.)

It's an exercise (just pick two bases, and associate their elements two by two) to show that one can always map two vector spaces V and W onto each other, if they have same dimension, by an invertible *linear* map (i.e., a map f such that $f(v + w) = f(v) + f(w)$ and $f(\lambda v) = \lambda f(v)$). So if one is using V to model some physics, W will do just as well. For instance, if the trajectory $t \to v(t)$ in V describes the evolution of a physical system, the trajectory $t \to f(v(t))$ in W provides an equivalent description. This is due to V and W being "of the same form", or as one says, *isomorphic*, via the *isomorphism* f.[5] From this point of view, there is only one *abstract* n-dimensional vector space, and this particular mathematical object we shall label V_n, for future reference.

Isomorphism doesn't mean that V_3 should be *identified* with $I\!R^3$. Indeed, V_3 and $I\!R^3$ can be put in one-to-one correspondence by a linear map: just select a basis $\{e_1, e_2, e_3\}$ in V_3, then the generic vector v can be expanded as $v = v^1 e_1 + v^2 e_2 + v^3 e_3$ and thus paired

reading, for better aspect on screens. In the process, "slanted" has become plain old italic. The layout has changed somewhat, too. But apart from typos, no substantial modifications have been done.]

[3] Not the same as *realizable*, of course, due to nonlinear effects: we know only too well that fields of arbitrary magnitude cannot be maintained in the hall. Note here how insidiously the modelling process gives status and credence to mathematical objects that may lack any counterpart in the real world: An electric field 10^{100} volts strong, for instance, is "conceivable", absurd as the very thought of one may be.

[4] It makes sense *from some vantage viewpoint*, of course. The technician in charge of the Van de Graaf may laugh off this "vector space" stuff as pedantic, whereas the person who simulates the experiment on a computer will see it as very natural.

[5] Linear maps are thus, among all possible maps between V and W, those which "preserve the linear structure". Needless to say, mathematicians have devised a language with which to discuss such abstract properties of abstractions. It's the theory of categories [LS], fondly nicknamed "abstract nonsense". A category regroups objects of similar structure, and *morphisms* are structure-preserving associations between pairs of such objects. Linear maps are the morphisms in the category of linear spaces. (A bit later, we'll meet affine maps, which are the morphisms in the category of affine spaces.)

with the triple $\{v^1, v^2, v^3\}$ of its *components* in this frame. But all this depends on the choice of basis, which is arbitrary. So there is no *canonical* way to associate V_3 and $I\!R^3$, which means, there is no unique, natural way to do it, that would stand out among all others for some good reason.

To say the same thing in different words, the vector v of V_3 can be *represented* by its components, but to say that v *is* the same as its triple of components would be going much too far. Unfortunately, pupils are trained all around the world to consider 3D vectors as triples of numbers, and when they begin to get the idea, one introduces the abstract idea of vector to them. There may be valid reasons for such pedagogy, but the (much sounder, I think) geometrical approach goes exactly the opposite: vectors are geometrical entities, which can be added, stretched, etc., and their Cartesian representation is only a useful computing device, by no means one that should always be used. (We all know examples of problems of geometry which can be solved more easily and more elegantly *without* coordinates than with them.)

As this theme will recur, I don't wish to hammer in the point right now, but just think about this: when solving a problem in electrotechnics, you don't pick just any system of coordinates (or "reference frame"), you carefully select one which is *adapted* to the device under study. So what makes the interest of coordinates is some peculiarity of the device, which makes some directions in space stand out among others. In the absence of such extra structure, there is no "canonical" way to select a set of coordinate axes, and imposing one would break the symmetry of space in an arbitrary way, devoid of physical justification.

Which prompts the question: What is this allusion to "symmetries of space" supposed to mean? This is where another basic structure, that of *group,* intervenes.

1.2 Affine space

Group is a simpler, more primitive, and hence more general structure than vector space. A vector space is already a group, an additive one: indeed, the operation $+$ admits of a *neutral* element (the vector 0, which is such that $v+0 = v$, for all v), and for each vector v, there is another one, namely $-v$, which yields 0 when added to v. Together with associativity (the fact that $u+(v+w) = (u+v)+w)$), these properties constitute the axioms for the group structure. Moreover a vector space is a *commutative* (or *Abelian)* group, meaning that $v + w = w + v$, something which is not part of the definition of groups in general. So a vector space can be construed as a set with *two* layers of structure: first, the one of Abelian group, conferred on it by vector addition; and a second layer, due to the introduction of scalar multiplication, with the properties required to make it compatible with addition. (This game of *peeling out* layers in a mathematical structure, in order to analyze it, we shall play recurringly.)

So if we forget about the multiplication by reals in V (thus depriving it of one of its structuring features) what remains is an Abelian group, called the associated group of *translations*. Why this name? It's a way to put emphasis on what a given vector v may *do* on other vectors, how it *acts* on them. To vector v, we may associate the map $w \to v+w$, called the "v-translation", and that we shall denote by T_v, so that $v + w$ is the same as $T_v(w)$, the *image* of w under the translation, or "v-translate" of w. Note that T_v is not a *linear* map of V to itself, because 0 does not map to 0. On the other hand, translations are not devoid of "linear" properties, since for instance, it is true that $T_v[(w + w')/2] = [T_v(w) + T_v(w')]/2$ (see Fig. 2—and note how one *must* divide by 2!), and a name will be useful for maps with this property: they are called "affine maps" of V to itself, or "affine transforms" on V.

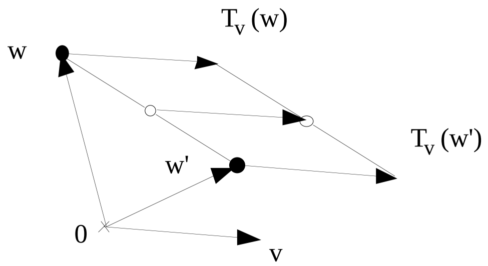

Figure 2. Translation by v, and its affine properties.

Obviously, the vector space structure is not required if one wants to define affine maps. All that is needed is a set, elements of which are now called *points,* not vectors, in which it makes sense to take the point midway between two points, or *barycenter* (and more generally, the barycenter of a given finite set of points to which weights of nonzero sum are assigned). Affine maps are then defined as maps which preserve barycenters. A set thus equipped with a notion of barycenter is called an *affine space*.

Given a vector space V, there is a way to build from it an associated affine space A, which is the same as V *as a set*, but differently structured. For this, we first define A as a set on which one can perform v-translations: Given a vector v of V and a point a of A, there is another point, denoted $T_v(a)$, which is the v-translate of a, and one postulates the obvious properties ($T_v(T_w(a)) = T_{v+w}(a)$, and so forth). One says that V "*acts* by translations" on A. Next, we assume that any two points of A can be connected by a v-translation, i.e., there is always some v such that $T_v(a) = b$, and that $T_v(a)$ always differs from a when $v \neq 0$. Nothing more natural, now, than writing $v = b - a$, or $T_v(a) = a + v$, so if one selects in A some particular point, denoted 0, one can make the identification between the point $T_v(0)$, alias $0 + v$, and the vector v. Not a *canonical* identification, of course!

Now the barycenter of a and b is the well defined $(b-a)/2$-translate of a, that is, the point $a+(b-a)/2$. *Affine maps*, from one affine space to another, are then defined, as already said, as those which preserve barycenters. (Affine *transforms* are affine maps from the space into itself. Those of them which are one-to-one, and thus invertible, form of course a group, called the *affine group*[6].) They preserve many other properties, as a consequence. For instance, alignment: three points "on the same line" (i.e., one of them is the barycenter of the other two, with adequate weights) are transformed to aligned points. Pairs of parallel lines transform into pairs of parallel lines, and so on. But distances or angles are *not* preserved. In fact, such notions simply don't make sense in affine space: to give them status, we shall have to introduce (but only later) another element of structure, called the *metric* of space. (Note right now, however, that *ratios* of distance between *aligned* points do make sense and are preserved, i.e., are "affine invariants", as one says.)

Everyday examples of affine transformations abound. If for instance, in a museum, you compare a painting with its catalogue reproduction, the two images in your visual field correspond by affine transform (Fig. 3), at least if they are small enough to allow the use of parallel perspective (the one where the eye is supposed to be at infinity).

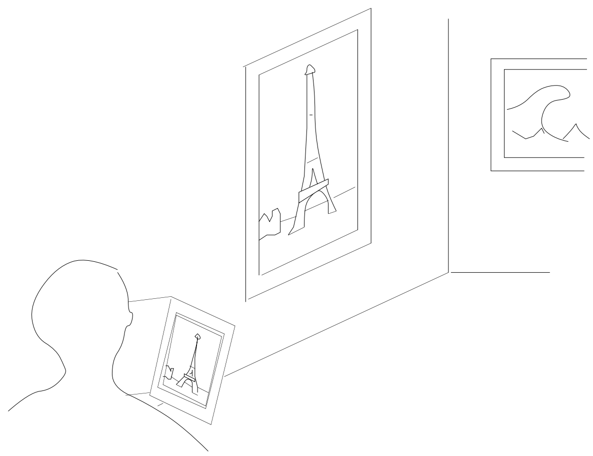

Figure 3. A common case of affine transform. (Note that "vanishing points" are at infinity in this rendering of the situation, so that parallel lines in the painting are seen as parallel lines of the views. More realistic perspective would defeat our purpose!)

To the abstract vector space V_3, it thus corresponds the abstract affine space A_3 (its dimension is, by definition, the dimension of the vector space). Informally, A_3 is what one gets when "forgetting where the origin was" in V_3. Conversely, selecting an origin in A_3 yields V_3, in a non-canonical identification. If one notices that selecting an origin for the space we live in is always an arbitrary move, A_3 emerges as a better model for ambient space[7] than V_3 and—a fortiori—$\mathbb{R}^3$.

In particular, the notion of vector field needs affine space, not only a vector space, to make sense. A *vector field* is a mapping from A_n to V_n. This mathematical entity is very apt to

6 At this stage, the reader may wish to examine the relation between this affine group, denoted GA_n in dimension n, and the more familiar linear group GL_n of linear transformations in V_n, which is isomorphic (via selection of a basis) to the group of $n \times n$ regular matrices. As a starting point, note that a linear transform on vectors induces an affine transform on points, and that, conversely, any affine transform can be described as the combination of such a special affine transform with a translation. Beware, this is more difficult than one might think.

7 Meaning, the space we live in. Later we shall use "ambient" in a more technical sense: it will refer to the encompassing space in the modelling, the one in which geometrical objects under consideration all live, a 3D affine space usually, but it may happen to be a manifold of any dimension.

model the (physical) notion of velocity field of, for instance, a mass of fluid: at each point x of A_3, fluid particles have a definite average velocity, represented by a vector $v(x)$ of V_3.

By the way, there is a name, *bound vector,*[8] $\{x, v(x)\}$ consisting of a point in A_n and a vector of V_n which one considers as assigned to this point. (One may also say "a vector at x".) Note that "bound" is *not* a qualifier, there: "bound vector" must be understood as a non-separable aggregate of words. Bound vectors are not vectors, one may even argue, because they do not form, taken together, a vector space: indeed, it makes no sense to add $\{x, v\}$ and $\{y, w\}$, unless $x = y$, or to multiply $\{x, v\}$ by λ. (They do form an affine space, though. Can you see it? First note that a bound vector $\{x, v\}$ can be construed as the pair of points $\{x, x + v\}$.) We are all familiar with the graphic convention according to which a family of bound vectors scattered on the page serves as a picture of a vector field.

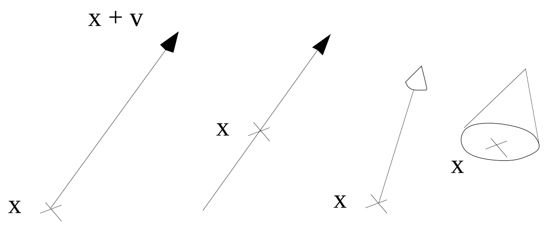

Figure 4. A few standard icons for bound vectors. For other examples, look carefully at plots of vector fields displayed by commercial software packages. As a rule, "three-dimensional" icons, like the two on the right, are to be preferred for 3D fields, and the more compact the icon, the better. (The norm of the vector, in the rightmost one, is rendered by the apparent volume of the cone.) The art of iconology, as applied to the visualization of fields, is still in its infancy. For some serious work in the area, see [C&], and [Tf] for general guidelines. (A reference list of works *not* to be imitated would exceed the size of this Journal.)

This is perhaps the right place for an aside, devoted to the notion of *icon* [Al]. Icons are drawings that stand for an abstract object, be it on our computer's screen or on a piece of paper. The most common icon for the bound vector $\{x, v\}$ is an arrow based at x with its tip at $x + v$ (Fig. 4). It's not *that* good a graphic convention, however, when it comes to visualize fields, because too long arrows tend to clutter in ugly tangles in regions where the field is large, symmetries that may exist are blurred, etc., thus such images often give a wrong idea of the overall field. The pair of points $\{x - v/2, x + v/2\}$ is often a better choice. One may also draw arrows differently, as suggested by Fig. 4.

1.3 Symmetries of physical space and of affine space

Now let's return to these alleged "symmetries" of space. Mathematically speaking, the symmetries of a structure are just its structure-preserving maps, so the symmetries of affine space are affine transforms, by definition. But what is at stake in the present discussion is something else: the ability of such notions to reflect symmetries *of the real world around us*. One of these symmetries is translational invariance: if you move this experiment hall fifty kilometers away, you will observe the same physics inside it. Humankind learned about this long ago, at least as regards *horizontal* translations. With Galileo and Newton, we realized that space was also invariant along the third dimension. Obvious changes in physical phenomena when one climbs up were attributed to the very presence of the Earth and of its gravitational field, but one accepted the idea that the laws of, say, celestial mechanics, would be the same a few light-years away, in any direction. So a prerequisite for all mathematical models of physical space would be translational invariance, and from this point of view, of course, affine space A_3 does qualify.

However, one may object, affine space is *too* symmetrical for the purpose, for one cannot pretend that physics is invariant with respect to scaling and shearing. But *some* aspects of physics are, as there exist experiments which can entirely be described using affine notions only. Figure 5 gives one (the idea comes from [Br], p. 100): The visitor of the science museum is watching a ball rolling along a gutter secured to the wall, and what we see, as in Fig. 3, is a photograph taken from infinity. Our view and his are different, but they correspond via some affine transform, which is enough to agree on our respective predictions: the ball will settle at the point of contact of the gutter with a line *parallel* to the wall's bottom line.

[8] Caution: Many physicists say "bound" and "free" where I say "free" and "bound", a usage that Burke [Bu] also endorses. But calling "free" a vector whose tail is attached to a point is more than I can swallow.

(The floor is supposed to be level, of course, but the wall need not be vertical, as far as it is plane.) Only affine notions are involved.

Admittedly, this is a very special case, and there are many physical events in which the symmetry of affine space is broken. Solid dynamics, for instance: rigid bodies you can translate and rotate, but not stretch or deform without altering their inner structure. Rigid bodies are so important in our existence that we need to be able to distinguish, among affine transforms, those which are "deformation-free". Mathematically, what is required for that is a *metric* structure.

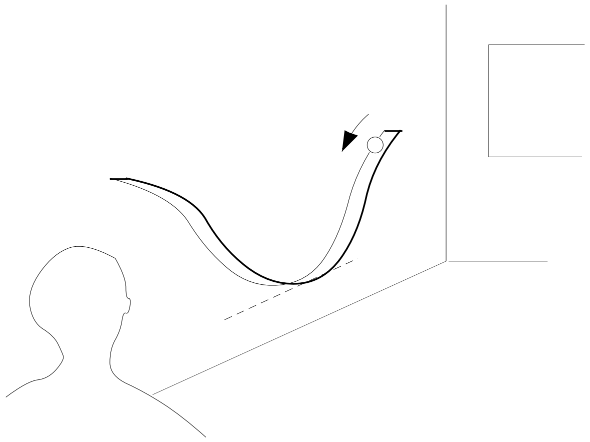

Figure 5. Where will the rolling ball eventually stop? In spite of our different perspectives, we agree with the visitor on that.

1.4 Metric

Metric is conferred onto a vector space V by endowing it with a *dot product*. The dot product $v \cdot w$ of two vectors v and w is a real number, and the correspondence $\{v, w\} \to v \cdot w$ is supposed to be linear with respect to both arguments, symmetrical (i.e., $v \cdot w = w \cdot v$) and most importantly, $v \cdot v > 0$ unless $v = 0$. Then, the square root $|v|$ of $v \cdot v$ is called the *norm* of v (more precisely, the *Euclidean* norm, as there are other kinds of norm—but we won't have to deal with them).

There are a lot of possible dot products on V_n. A way to get them all is to select some frame $\{e_1, e_2, \ldots, e_n\}$ and to set $v \cdot w = \sum_{i,j} g_{ij} v^i w^j$, where the *metric coefficients* g_{ij} are the entries of a square *strictly positive definite* symmetric matrix. As soon as we have adopted a dot product, notions of orthogonality and angle begin to make sense. We also know what "distance" means, speaking of two points x and y of the associated affine space: it's the number $d(x, y) = |x - y|$, of which one immediately sees it satisfies all properties required of a distance ($d(x, y) = d(y, x) > 0$, unless $x = y$, and the triangle inequality). This turns A_n into a *metric space* (any space equipped with such a distance function), with a bonus: the metric is *compatible with the affine structure*, which means that it's invariant by translations, $d(x + v, y + v) = d(x, y)$. Translations are thus *isometries*, i.e., transforms that preserve distance. It's relatively easy to show that isometries in A_n must be affine transforms (let's not feel forced to do it here). But the converse is not true. Those affine transforms that do preserve distances are called *displacements*. They form of course a group, smaller than GA_n. Among them, those that fix at least one point are called *orthogonal transforms*. They include *rotations* (but rotations have another property: they "preserve orientation", a notion we shall soon discuss and criticize), and *mirror reflections* (transforms $v \to v - 2\, u{\cdot}v\, u$, where u is a *unit vector* $(|u| = 1)$. Orthogonal transforms which fix a given point form what crystallographers call a *point group*.[9]

Granted that a metric structure is necessary to correctly model solid dynamics, should we commit ourselves to a metric when only electromagnetism is involved? Definitely, yes. As Fig. 6 should suggest, there are optical experiments which cannot be described in exclusively affine terms, the way a ray of light bounces off a mirror, for example.

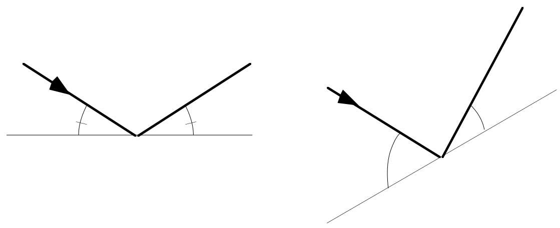

Figure 6. Equality of angles in light-ray reflection is not an affine notion. (Symmetry of the ray and its reflection with respect to the normal is an affine notion, but it's the concept of *normal*, now, which is not an affine one!)

What is very exciting, however, is that all aspects of Maxwell's theory are not alike in this

[9] The reader may wish to check that such a group is isomorphic to the group of orthogonal matrices of order n. A *space* group, in contrast to *point* group, is a subgroup of displacements which contains translations in all three directions.

respect. Some are affine invariant, some require a metric. As we shall little by little discover, both Faraday's law and Ampère's theorem *can* be edicted using only affine notions. On seeing box b of Fig. 1, this statement seems utterly unlikely, doesn't it? For the very definition of curl *does* involve a metric (just try to change the scale of length, for instance). Yet it's true, as we shall see. But this truth is hidden by having represented the physical entities forming the (physical) EM field by these mathematical entities, the vector fields **E**, **H**, **D**, and **B**.

On the other hand, the constitutive laws $\mathbf{B} = \mu\mathbf{H}$ and $\mathbf{D} = \epsilon\mathbf{E}$ in box b are metric-dependent. This is where the metric structure of space intervenes[10] in the laws of electromagnetism. The modern notation of box c neatly makes the distinction, as metric is concentrated, as we'll see later, in the "star-operator" of the middle equation.

In fact, this differential geometric notation is even more general, for the equations $dF = 0$ and $dG = J$ do not depend on all aspects of the affine structure. (Let's say rapidly, though we don't have the technical equipment for such issues at this stage, that "their invariance goup is (much) larger" than the affine group GA_4.) Indeed, Maxwell equations continue to make sense, and to be physically relevant, in situations where the underlying space does *not* possess the symmetries we are used to, as for instance when investigating the magneto-hydrodynamics of a dense star, where space is "warped", as one knows, according to General Relativity. Although physicists, obviously, need to deal with such situations, engineers don't (well, not yet . . .), so the choice of metricized affine space as the framework in which to do our modelling is a reasonable one.

1.5 Orientation

Still, a last element of structure is lacking, which one cannot do without in electrodynamics: orientation of space. Among elements of the framework we are building, it certainly is the most difficult one to discuss, and the source of endless difficulties experienced by students, and not only them, when dealing with fields.

Orientation, like metric, is an element of structure that one may lay over a vector space. These are independent structures. One may have metric without orientation, and the other way around. So here, we assume a given vector space of dimension n, but no dot product.

Consider two frames in V_n , say $\{e_i : i = 1, \ldots, n\}$ and $\{f_j : j = 1, \ldots, n\}$. One may express the e_is as linear combinations of the f_js, hence a "transition matrix" T such that $e_i = \sum_j T_i^j f_j$. As T is regular, its determinant has a definite sign, $+$ or $-$. We say that $\{e_i\}$ and $\{f_j\}$ have *the same orientation* if the sign is $+$, *opposite* orientations if the sign is $-$. (Obviously, the sign is the same if one expresses the f_js in the $\{e_i\}$-frame.) This defines two classes of frames, two of them belonging to the same class if they have same orientation. An *oriented vector space* is a composite mathematical object, a pair, which consists of (1) a (finite dimensional) vector space, (2) *one* of its two orientation classes. So for each vector space, there are *two* oriented vector spaces, with opposite orientations, which can be associated with it. To *orient* V_n consists in making a choice between these two possibilities, that is, designating a distinguished class of frames. It's convenient to name this class Or, and the other one $-Or$. Frames of Or will then be called *direct* frames with respect to this orientation), and those of $-Or$, *skew* frames. (One also says "even" and "odd", hence the notion of "parity" of a frame, which is just the class it belongs to.) The two possible oriented spaces are thus $\{V_n, Or\}$ and $\{V_n, -Or\}$, which we shall abbreviate as ${}^+V_n$ and ${}^-V_n$. (Of course, if one selects ${}^-V_n$ as the oriented space in which to work, then the skew frames are those of Or.)

Take good note that, once a vector space has been oriented, there are direct frames and skew frames, but there is no such thing as direct or skew vectors, except, one may concede, if $n = 1$. A vector is a vector is a vector, and does not become a new object just because the space it belongs to has been oriented! This remark will be important later in our discussion of polar and axial vectors.

Remark. It's all right to consider an oriented vector space as a pair consisting of (1) a space,

[10] And though it's much too early, one can't resist the urge to confirm what the reader may already be suspecting: *things go the other way.* It's the constitutive laws of electromagnetism that give space its metric. After all, don't we make our geodetic surveys with light rays?

(2) *one* of its frames, provided the frame thus privileged serves no other purpose than fixing the orientation class. Pupils asked to "orient the figure" are, unfortunately, often confused by that, for they tend to believe that this is the same as selecting coordinate axes. Not so. Orienting the paper sheet ($n = 2$) means deciding on a "direct" sense of rotation (anticlockwise, most often), but one is not committed to definite axes by that. Orienting space ($n = 3$) means deciding which helices are direct or skew. As one knows, the usual convention for orienting 3D space is the "corkscrew rule", which makes most helices of the real world (shells, staircases, ...) direct, or as one also says, right-handed. ♢

An affine space, now, is oriented by orienting its associate vector space: a "bound frame" at x in A_n , i.e., a set of n independent vectors at x, is direct or skew if the n vectors form a direct or a skew frame in V_n. Hence two new structures: ${}^+A_n$ and ${}^-A_n$.

Vector subspaces of a given vector space (or affine subspaces of an affine space) can have their own orientation. Orienting a line, in particular, means selecting a vector parallel to it, called the *director* (vector) of the line, which points in what is then, conventionally, the "forward" direction along this line. Note that such orientations of different subspaces are a priori unrelated. Orienting 3D space by the corkscrew rule, for instance, does not imply any orientation in a given plane. Still, the standard orientation of space and of a horizontal plane do match, obviously. How come? Because the vertical direction also is oriented, bottom-up. So, if space is oriented, and if some privileged direction in space is oriented, planes that are "transverse" to this direction (meaning, the intersection reduces to a single point) inherit an orientation, as follows: to know whether a frame in the plane is direct or skew, just append it to the director of the line (i.e., place the latter ahead of the list of frame vectors), and check whether the spatial frame thus obtained is direct or skew.

The recipe can be generalized to all dimensions, so let's introduce a convenient terminology. We say that two subspaces U and W of V are *complementary* if their *span* is all V (i.e., if any v in V can be decomposed as $v = u + w$, with u in U and w in W) and if they are *transverse* ($U \cap W = \{0\}$, which makes the decomposition unique). Now (Fig. 7), we say that U has an *external* or *outer* orientation if an orientation is provided for one of its complements, W say. (For contrast and clarity, we shall call *inner* orientation what was simply "orientation" up to this point.) These notions (which one can trace back to [VW], cf. [VD] and [Sc]) pass to affine subspaces of an affine space the obvious way.

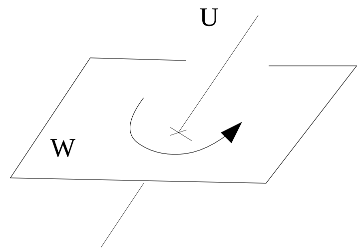

Figure 7. Externally orienting a line U by orienting a plane W transverse to it.

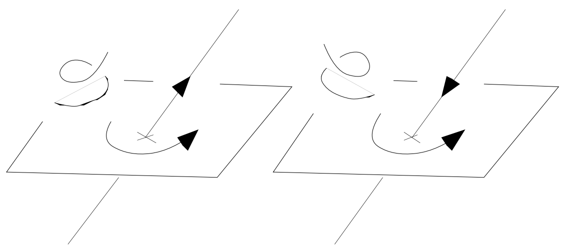

Figure 8. How an externally oriented line acquires inner orientation, depending on the orientation of ambient space. Alternative interpretation: if one knows both orientations, inner and outer, for a line, one knows the ambient orientation. The drawing on the left, then, can be understood as an explanation of *Ampère's rule*.

It's clear that if the encompassing space V itself is oriented, then an outer orientation of U gives it an inner orientation: to know the orientation class of a frame in U, append it to a direct frame of W, thus obtaining a frame in V, and look to which class the latter belongs. But two possible orientations for V make *two* ways to do that, so outer orientation and inner orientation are different, as are their intuitive meanings. For instance, inner orienting a line means distinguishing "forward" and "backward" directions along it. But outer orienting the line, that is to say, inner orienting a transverse plane, amounts to make a choice between the two ways to "turn around" the line. If ambient space is oriented, the "direct" way to turn around a line implies a way to go "forward" along it (see Fig. 8, and note how a play on icons advantageously substitutes for all this stilted prose!). Similarly, outer orienting a plane means specifying a "crossing direction" through it.

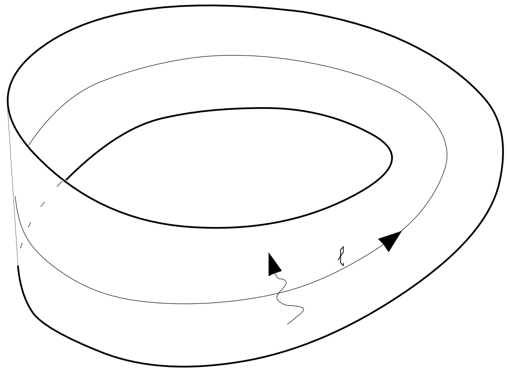

Figure 9. Möbius band, not orientable. (To prove this "experimentally", make such a band, lay on it the drawing of a frame, that will be moved around the band, thus returning at the starting point in inverted position. Be careful to use *transparent* material for both the ribbon and the frame, otherwise, your parlour-trick will fail on its nose.) As the middle line l does not separate two regions, no global crossing direction can be defined for it, so it has no outer orientation with respect to the band.

Curved lines also can be internally and externally oriented. The case of surfaces is a bit more complex. Orienting a surface means orienting all of its tangent planes at all points, in a "consistent" way. For neighboring points, tangent spaces are different, but close enough to have a common transversal. Orientations at these points are consistent if they give the same outer orientation to this transversal, which can always be achieved. But though this ensures consistent orientation locally, it may not be possible to maintain such consistency all over, as this all-time star of mathematical popularizations, the Möbius band, testifies (Fig. 9).

On the other hand, surfaces which enclose a volume can be oriented: "going inside out" defines a consistent crossing direction. This is outer orientation, from which inner orientation stems, if the ambient space is oriented.

Remark. While inner orientability is an intrinsic property, outer orientability always refers to some ambient space. It makes sense, for instance, to speak of outer-orienting a line traced on a surface: this means, as above, defining a consistent crossing direction, from surface points on one side of the line to points on the other side. As Fig. 9 shows, this may not be possible for some lines when the encompassing surface is non-orientable. This demonstrates how different the two notions of orientation can be. $\diamondsuit$

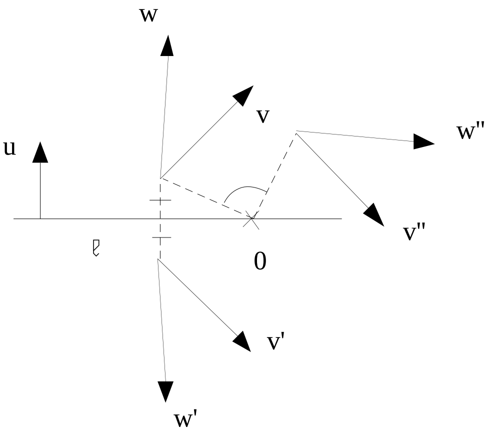

Figure 10. A skew transform (mirror reflection about line l, with unit vector u), and a direct one (rotation around 0).

Before leaving orientation, we need to broach this dangerously vague notion that some geometrical transforms could either "preserve" or "reverse orientation". What is meant by that is their effect on *frames*. Apply an invertible affine transform to n bound vectors forming a frame, you get another bound frame (at another point, in general, cf. Fig. 10). The two frames belong to the same class or they don't. Hence two classes of transforms: the *direct* ones (like rotations), for which a frame and its image belong to the same orientation class, and the *skew* ones (the other way round), like mirror reflections. *(Central symmetry,* i.e., the affine map $x \to -x$, can be direct or skew, depending on the dimension.) One also says "parity preserving" and "parity reversing" transforms.[11] Note that the orientation class Or, as a whole, is mapped to $-Or$ by an odd transform, but this cannot by any means "change the orientation of space", that is to say, our earlier commitment to Or as the class of direct frames!

1.6 Oriented Euclidean space

We are now in possession of a framework in which to model electrical phenomena: *oriented Euclidean three-dimensional affine space*, that will be denoted E_3 (and $^+E_3$ when it will be

[11] The notion applies to more general point-to-point transforms than those of GA_n; but the parity of such a transform, then, is only locally defined, and may not be the same at all points.

felt necessary to remind about orientation). It's A_3 coated with two layers of structure: a dot product and an orientation.

Dimension 3 has this in particular[12] that one can define a new operation, the cross product: Given two vectors u and v, the cross product $u \times v$ is a vector orthogonal to both of them, of squared length $|u|^2|v|^2 - (u \cdot v)^2$, and such that the frame $\{u, v, u \times v\}$ be direct. The very notion, therefore, does not make sense without a metric *and* an orientation. To keep oneself aware of that, one might "decorate" the symbol $\times$, like this: ${}^{\cdot}\!\times^{+}$ or ${}^{\cdot}\!\times^{-}$. Notice that $u \;{}^{\cdot}\!\times^{+}\; v = -u \;{}^{\cdot}\!\times^{-}\; v$, which clearly explains what is meant when one says that "$\times$ is sensitive to orientation". This would be, however, the only advantage of such heavy notation, which one can't seriously propose.

Remark. If $\bullet$ is a new dot product, one has $u \bullet v = Lu \cdot Lv$, where L is some linear map. One should then be able to express $u \;{}^{\bullet}\!\times^{+}\; v$, the cross product associated with this new metric, in terms of u, v, L, and ${}^{\cdot}\!\times^{+}$ (or perhaps, ${}^{\cdot}\!\times^{-}$). Would you care to try it? It's not so easy an exercise. $\diamondsuit$

Another kind of "sensitivity to orientation" is demonstrated by the following fact: if T is an affine transform, $T(u \times v) = \pm Tu \times Tv$, the sign depending on the parity of T (it's $-$ for a mirror reflection). This contrasts to what happens with respect to vector addition, since one has $T(u + v) = Tu + Tv$, unconditionally.

We can briefly summarize all that by saying that the cross product operation belongs to the structure of *oriented* 3D Euclidean space.

This is true of other operations, notoriously the curl operator. Let's not try to define the curl the "elegant" way, without coordinates, because it's one of these cases where one is better off using them. Having adopted a direct orthonormal Cartesian system of basis vectors and axes, start from a smooth vector field u, take the curl the usual way, and just check that the field rot$u$ thus obtained would have been the same with another system of such axes, which is easy by invoking the Stokes theorem. It's then clear (Fig. 11) that rot$(Mu) = -M(\text{rot}\, u)$, if M is the mirror reflection with respect to a plane m.

[12] One might generalize to $n-1$ vectors in E_n , but this is not usual.

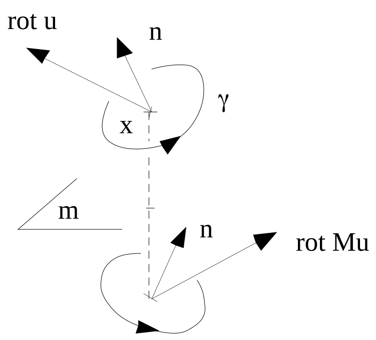

Figure 11. Applying Stokes' theorem to a small patch of surface around x, rimmed by γ, and to its mirror image, to see that rot$(Mu) = -M(\text{rot}\, u)$.

The other differential operations, grad and div, are less capricious. We'll discuss grad at some length next time, but it's obviously immune to orientation diseases. As for div, it's even simpler, as the metric structure is irrelevant in this case: the operator div belongs to the affine structure. (Think of v as the velocity field of some compressible fluid. Then div$v$ expresses the *rate* of change of the volume along the flow, which is an affine concept.)

This concludes our survey of "space", as a framework for modelling: We shall work in E_3, "oriented Euclidean 3D space", while being well aware of the "multilayered" character of this structure. Logically, we should discuss "time" as well, but having no ambition to address Relativity here, we shall be content to consider time as a parameter, which is all right if all phenomena are referred to the previous space E_3. The next part will introduce new geometrical objects: axial vectors, covectors (instead of vectors), and differential forms (instead of vector fields), whose introduction will be motivated by an analysis of the Lorentz force exerted on a moving charge.

References

[Al] S.L. Altmann: **Icons and Symmetries,** Clarendon Press (Oxford), 1992.

[Br] W.L. Burke: **Spacetime, Geometry, Cosmology**, University Science Books (Mill Valley, CA 94941), 1980.

[Bu] W.L. Burke: **Applied Differential Geometry,** Cambridge University Press (Cambridge, U.K.), 1985.

[Cr] M.J. Crowe: **A History of Vector Analysis**, University of Notre Dame Press, 1967 (Dover edition, New York, 1985).

[C&] V. Čingoski, M. Ichinose, K. Kaneda, H. Yamashita: “Analytical Calculation of Magnetic Flux Lines in 3-D Space”, **IEEE Trans., MAG-30**, 5 (1994), pp. 2912-5.

[LS] F.W. Lawvere, S.H. Schanuel: **Conceptual Mathematics,** A first introduction to categories, Cambridge U.P. (Cambridge), 1997.

[Ma] J.C. Maxwell: **A Treatise on Electricity and Magnetism**, Clarendon Press, 3d ed., 1891 (Dover edition, New York, 1954).

[Mi] C.W. Misner, K.S. Thorne, J.A. Wheeler: **Gravitation**, Freeman (New York), 1973.

[Sc] J.A. Schouten: **Tensor analysis for physicists**, Dover (New York), 1989.

[Sp] R.J. Stephenson: “Development of vector analysis from quaternions”, **Am. J. Phys., 34**, 2 (1966), pp. 194-201.

[Tf] E.R. Tufte: **The Visual Display of Quantitative Information,** Graphics Press (Box 430, Cheshire, Conn. 06410), 1983.

[VD] D. Van Dantzig: “On the geometrical representation of elementary physical objects and the relations between geometry and physics”, **Nieuw. Archief vor Wiskunde, 3**, 2 (1954), pp. 73-89.

[VW] O. Veblen, J.H.C. Whitehead, **The foundations of differential geometry**, Cambridge, 1932.