

HANDOUT #13: INTRODUCTION TO THE ROTATION GROUP

## 1 Lie groups

As mentioned in Handout #11, the general framework for studying symmetry in mathematics and physics is **group theory**. In your algebra class you have studied many examples of finite groups (e.g. the symmetric group and its subgroups) and some examples of infinite discrete groups (e.g. the integers  $\mathbb{Z}$ ). But in physics the most important symmetries are those associated to *continuous groups* (i.e. groups parametrized by one or more real parameters), since these are the ones that give rise, via Noether’s theorem, to conservation laws. Continuous groups are also known as **Lie groups**, in honor of the Norwegian mathematician Sophus Lie (1842–1899); they play a key role in differential geometry, quantum mechanics, and many other areas of mathematics and physics. In this handout I would like to give you an elementary introduction to the theory of Lie groups, focusing on the classical matrix groups and then concentrating on the most important one for our purposes, namely the group of rotations of three-dimensional space.

A precise definition of “Lie group” requires some preliminary concepts from differential geometry, notably the notion of a *manifold*. Instead of doing this, I will simply show you some concrete examples of Lie groups; all of my examples will be groups of matrices, with the group operation being matrix multiplication.

### 1.1 The general linear group

The set of all  $n \times n$  matrices (with real entries) does not form a group with respect to matrix multiplication, because not all matrices have multiplicative inverses. But if we restrict ourselves to *invertible* matrices, then we do have a group, since the product of two invertible matrices is invertible:  $(AB)^{-1} = B^{-1}A^{-1}$ . We therefore define the **general linear group**  $GL(n)$  to be the set of all invertible  $n \times n$  matrices:

$$GL(n) = \{A \in \mathbb{R}^{n \times n} : A \text{ is invertible}\} . \quad (1)$$

Since a matrix is invertible if and only if its determinant is nonzero, we can also write

$$GL(n) = \{A \in \mathbb{R}^{n \times n} : \det A \neq 0\} . \quad (2)$$

Therefore,  $GL(n)$ , considered as a set, is simply the  $n^2$ -dimensional vector space  $\mathbb{R}^{n \times n}$  with the subset  $\{A : \det A = 0\}$  removed. Since the determinant is a polynomial function of the matrix elements  $a_{11}, a_{12}, \dots, a_{nn}$ , it follows that the set  $\{A : \det A = 0\}$  is an “algebraic subvariety of codimension 1” in  $\mathbb{R}^{n \times n}$ . [Codimension 1 because it is specified by *one* equation.]

Geometrically it is a *hypersurface* in  $\mathbb{R}^{n \times n}$  (“hypersurface” is a synonym of “submanifold of codimension 1”). Fancy language aside, the important point is that “nearly all”  $n \times n$  matrices are invertible; only a lower-dimensional subset of them are noninvertible. So  $GL(n)$  is a Lie group of the full dimension  $n^2$ .<sup>1</sup>

The  $n \times n$  matrices are in one-to-one correspondence with the linear maps from  $\mathbb{R}^n$  to itself: namely, the matrix  $A$  induces the linear map  $x \mapsto Ax$ . Under this correspondence, the matrices  $A \in GL(n)$  correspond to the *automorphisms* of the  $n$ -dimensional real vector space  $\mathbb{R}^n$ : that is, they are the maps from  $\mathbb{R}^n$  to itself that preserve its vector-space structure (namely, are linear and invertible). The group  $GL(n)$  can thus be regarded as the group of automorphisms of the  $n$ -dimensional real vector space  $\mathbb{R}^n$ .

All of the remaining groups to be introduced here will be subgroups of  $GL(n)$ .

**Remark.** Sometimes we write  $GL(n, \mathbb{R})$  to stress that these are matrices whose entries are real numbers. One can also study the group  $GL(n, \mathbb{C})$  of complex matrices, or more generally the group  $GL(n, F)$  for any field  $F$ .

## 1.2 The special linear group

The determinant of an  $n \times n$  matrix has an important geometric meaning: namely,  $\det A$  is the factor by which volumes are scaled under the transformation  $x \mapsto Ax$  of  $\mathbb{R}^n$  into itself. (More precisely,  $|\det A|$  is the factor by which volumes are scaled, while the sign of  $\det A$  tells us whether orientations are preserved or reversed.) This suggests that we might consider the subgroup of  $GL(n)$  consisting of matrices that preserve volume. We therefore define the **special linear group**  $SL(n)$  by

$$SL(n) = \{A \in GL(n) : \det A = 1\}. \quad (3)$$

Since  $\det(AB) = (\det A)(\det B)$ , it follows that  $SL(n)$  really is a group; it is a subgroup of  $GL(n)$ . Since  $SL(n)$  consists of the matrices  $A$  satisfying *one* equation  $\det A = 1$ , it is a subgroup of codimension 1. Therefore,  $SL(n)$  is a Lie group of dimension  $n^2 - 1$ ; it will be parametrized (at least locally) by  $n^2 - 1$  independent real parameters.

**Remark.** Once again we write  $SL(n, \mathbb{R})$  to stress that these are matrices whose entries are real numbers, since one can also study  $SL(n, \mathbb{C})$  and  $SL(n, F)$  for any field  $F$ , or even  $SL(n, R)$  for any commutative ring  $R$ . In fact, the group  $SL(2, \mathbb{Z})$  and its subgroups play an important role in complex analysis and analytic number theory. (Note, however, that  $SL(2, \mathbb{Z})$  is a discrete group, not a Lie group.)

## 1.3 The orthogonal group and special orthogonal group

Now consider  $\mathbb{R}^n$  equipped with the Euclidean inner product  $x \cdot y = \sum_{i=1}^n x_i y_i$  [where  $x = (x_1, \dots, x_n)$  and  $y = (y_1, \dots, y_n)$ ] and hence with the Euclidean norm  $\|x\| = (x \cdot x)^{1/2}$ . Which linear transformations  $x \mapsto Ax$  preserve these structures? You learned the answer in your linear algebra course:

---

<sup>1</sup>Another way of saying this is: for any invertible matrix  $A$ , there exists a neighborhood of  $A$  in  $\mathbb{R}^{n \times n}$  in which all the matrices are invertible.

**Lemma 1** Let  $A$  be an  $n \times n$  matrix with real entries, and consider the linear transformation of  $\mathbb{R}^n$  defined by  $x \mapsto Ax$ . Then the following are equivalent:

- (a) This transformation preserves the Euclidean inner product, i.e.  $(Ax) \cdot (Ay) = x \cdot y$  for all  $x, y \in \mathbb{R}^n$ .
- (b) This transformation preserves the Euclidean norm, i.e.  $\|Ax\| = \|x\|$  for all  $x \in \mathbb{R}^n$ .
- (c) The matrix  $A$  is an **orthogonal matrix**, i.e.  $A^T A = I$ .

For completeness let me remind you of the proof:

PROOF. Let  $x$  and  $y$  be arbitrary vectors in  $\mathbb{R}^n$ , and let us compute

$$(Ax) \cdot (Ay) = \sum_{i=1}^n (Ax)_i (Ay)_i \quad (4a)$$

$$= \sum_{i,j,k=1}^n A_{ij} x_j A_{ik} y_k \quad (4b)$$

$$= \sum_{i,j,k=1}^n x_j (A^T)_{ji} A_{ik} y_k \quad (4c)$$

$$= \sum_{j,k=1}^n x_j (A^T A)_{jk} y_k. \quad (4d)$$

So if  $A^T A = I$ , or in other words  $(A^T A)_{jk} = \delta_{jk}$ , we have  $(Ax) \cdot (Ay) = x \cdot y$ . So (c) implies (a).

For the converse, let  $x = \mathbf{e}_\ell$  (the vector having a 1 in the  $\ell$ th entry and zeros in all other entries) and  $y = \mathbf{e}_m$ . Then  $x \cdot y = \delta_{\ell m}$ . And the formula above shows that  $(Ax) \cdot (Ay) = (A^T A)_{\ell m}$ . These two are equal for all choices of  $\ell, m$  only if  $A^T A = I$ . So (a) implies (c).

Now (a) trivially implies (b), by setting  $y = x$ . To prove the converse, we observe that

$$x \cdot y = \frac{1}{2} \left[ (x+y) \cdot (x+y) - x \cdot x - y \cdot y \right] \quad (5)$$

and

$$(Ax) \cdot (Ay) = \frac{1}{2} \left[ (A(x+y)) \cdot (A(x+y)) - (Ax) \cdot (Ax) - (Ay) \cdot (Ay) \right]. \quad (6)$$

It follows from these two formulae that (b) implies (a) [please make sure you understand why].  $\square$

**Remarks.** 1. The operation used in the proof of (b)  $\implies$  (a) is called **polarization**: it is a general procedure that allows one to recover a symmetric bilinear form from the quadratic form it generates. An alternative version of the polarization formula is

$$x \cdot y = \frac{1}{4} \left[ (x+y) \cdot (x+y) - (x-y) \cdot (x-y) \right] \quad (7)$$

and the analogous thing with  $A$ .

2. In your linear algebra course you learned that for a finite square matrix  $A$ , the following are equivalent:

- (a)  $A$  has a left inverse (namely, a matrix  $B$  satisfying  $BA = I$ ).
- (b)  $A$  has a right inverse (namely, a matrix  $B$  satisfying  $AB = I$ ).
- (c)  $\det A \neq 0$ .

and that when (a)–(c) hold, the left and right inverses are *unique* and *equal*. It follows from this that  $A^T A = I$  is equivalent to  $AA^T = I$ ; therefore, we can define “orthogonal matrix” by any one of the three equivalent statements:

- (i)  $A^T A = I$
- (ii)  $AA^T = I$
- (iii)  $A^T A = AA^T = I$

**Warning:** The equivalence (a) $\iff$ (b) is *not* true for rectangular matrices, or for infinite matrices, or for operators on infinite-dimensional vector spaces.  $\square$

In view of the preceding Lemma, we define the **orthogonal group**  $O(n)$  by

$$O(n) = \{A \in GL(n) : A^T A = I\} . \quad (8)$$

[You should verify for yourself that  $A, B \in O(n)$  imply  $AB \in O(n)$  and  $A^{-1} \in O(n)$ , so that  $O(n)$  is indeed a group; it is a subgroup of  $GL(n)$ .] The group  $O(n)$  is the group of automorphisms of  $n$ -dimensional Euclidean space (i.e.,  $\mathbb{R}^n$  equipped with the Euclidean inner product).

Taking determinants of the defining equation  $A^T A = I$  and remembering that

$$\det(A^T) = \det A \quad \text{and} \quad \det(AB) = (\det A)(\det B) , \quad (9)$$

we see that a matrix  $A \in O(n)$  must satisfy  $(\det A)^2 = 1$ , or in other words

$$\det A = \pm 1 . \quad (10)$$

Both signs are possible. Consider, for instance,  $n \times n$  diagonal matrices with 1’s and  $-1$ ’s on the diagonal. All such matrices are orthogonal, hence belong to  $O(n)$ . If the number of entries  $-1$  is even, then  $\det A = +1$ ; but if the number of entries  $-1$  is odd, then  $\det A = -1$ .

Topologically, the group  $O(n)$  is the disjoint union of two connected components: the orthogonal matrices with determinant  $+1$  (this is the connected component containing the identity matrix), and the orthogonal matrices with determinant  $-1$ . (These two subsets must “stay away from” each other in  $\mathbb{R}^{n \times n}$ , because the determinant is a continuous function.)

The orthogonal matrices with  $\det A = +1$  form a subgroup of  $O(n)$  [why?]; it is called the **special orthogonal group**  $SO(n)$ :

$$SO(n) = \{A \in O(n) : \det A = +1\} = O(n) \cap SL(n) . \quad (11)$$

It is also called the **rotation group**, because the elements of  $SO(n)$  can be regarded as rotations. (We will see this explicitly for  $n = 2$  and  $n = 3$ .) By contrast, some of the matrices in  $O(n)$  with determinant  $-1$  can be regarded as **reflections**.

What is the dimension of the Lie groups  $O(n)$  and  $SO(n)$ ? The defining equation  $A^T A = I$  naively has  $n^2$  entries. But it is important to note that the matrix  $A^T A$  is *symmetric* (no matter what  $A$  is); therefore, some of the  $n^2$  equations in  $A^T A = I$  are redundant (namely, the  $ij$  equation says the same thing as the  $ji$  equation). So the non-redundant equations are, for instance, those on the diagonal and those above the diagonal (i.e.  $i \leq j$ ); there are  $n(n + 1)/2$  of these [you should check this carefully]. Therefore, the dimension of the Lie groups  $O(n)$  and  $SO(n)$  is

$$n^2 - \frac{n(n + 1)}{2} = \frac{n(n - 1)}{2}. \quad (12)$$

In particular, the rotation group in dimension  $n = 2$  is one-dimensional (it is parametrized by a single angle), and the rotation group in dimension  $n = 3$  is three-dimensional (it can be parametrized by three angles, as we shall later see explicitly).

## 1.4 The pseudo-orthogonal groups

Here is a generalization of the orthogonal groups  $O(n)$  that plays an important role in special relativity, among other applications.

Note first that the defining equation  $A^T A = I$  of the orthogonal group can trivially be rewritten as  $A^T I A = I$ . That is, the orthogonal transformations are those that preserve the symmetric bilinear form represented by the matrix  $I$ , which is  $\langle x, y \rangle = x^T I y = x \cdot y$ . (This is precisely the content of the Lemma of the preceding section.)

So it might be worth generalizing this to consider matrices that preserve the symmetric bilinear form represented by some other symmetric matrix. For instance, we can consider the diagonal matrix  $I_{p,q}$  whose diagonal entries are  $p$   $+1$ 's followed by  $q$   $-1$ 's (where  $n = p + q$ ). The associated symmetric bilinear form is

$$\langle x, y \rangle = x^T I_{p,q} y = \sum_{i=1}^p x_i y_i - \sum_{i=p+1}^{p+q} x_i y_i. \quad (13)$$

We then define the **pseudo-orthogonal group**  $O(p, q)$  by

$$O(p, q) = \{A \in GL(p + q) : A^T I_{p,q} A = I_{p,q}\}. \quad (14)$$

In particular, the pseudo-orthogonal group  $O(3, 1)$  is called the **Lorentz group**; it plays a central role in special relativity. (The numbers 3 and 1 arise because we live in a universe having three space dimensions and one time dimension.)

It is easy to see, just as with the orthogonal groups, that every matrix  $A \in O(p, q)$  has determinant  $\pm 1$  and that both signs are possible. We therefore define also the **special pseudo-orthogonal group**  $SO(p, q)$  by

$$SO(p, q) = \{A \in O(p, q) : \det A = +1\} = O(p, q) \cap SL(p + q). \quad (15)$$

The pseudo-orthogonal groups  $O(p, q)$  and  $SO(p, q)$  have dimension  $n(n - 1)/2$  [where  $n = p + q$ ], just like the orthogonal group.

**Remark.** One might ask whether there are other symmetric bilinear forms worth considering, beyond  $I_{p,q}$ . The answer is *no*, because **Sylvester’s law of inertia** guarantees that every nondegenerate symmetric bilinear form can be represented by  $I_{p,q}$  in a suitably chosen basis. More precisely, let us say that the  $n \times n$  matrices  $A$  and  $B$  are *congruent* if there exists an invertible matrix  $S$  such that  $B = SAS^T$  (and hence also  $A = S^{-1}B(S^{-1})^T$ ). Then Sylvester’s theorem asserts that any symmetric matrix  $A$  is congruent to a diagonal matrix having all its diagonal entries equal to  $+1$ ,  $-1$  or  $0$ , and that moreover the number of diagonal entries of each kind is an invariant of  $A$ , i.e. it does not depend on the matrix  $S$  being used. This invariant is called the **inertia**  $(n_+, n_-, n_0)$  of the matrix  $A$ . The *nondegenerate* symmetric bilinear forms are the ones with  $n_0 = 0$ ; that is, they are congruent to some matrix  $I_{p,q}$  (where  $p = n_+$  and  $q = n_-$ ).

## 1.5 The symplectic group

We now come to the least intuitive of the “classical matrix groups”, namely the symplectic group. But I would be remiss if I ignored it, because it plays a central role in Hamiltonian mechanics.

We follow the same reasoning as for the pseudo-orthogonal groups, but instead of looking for matrices that preserve a symmetric bilinear form, we look for matrices that preserve an *antisymmetric* bilinear form. That is, we fix an antisymmetric matrix  $J$  and we look for matrices  $A$  satisfying  $A^T J A = J$ . Now, nondegenerate antisymmetric matrices exist only in *even* dimension, say dimension  $2n$ .<sup>2</sup> The canonical example of a nondegenerate antisymmetric matrix is precisely the matrix  $\Omega$  that we introduced in Handout #12 to represent the structure of Hamiltonian phase space. Namely,  $\Omega$  is the  $2n \times 2n$  matrix whose  $n \times n$  blocks look like

$$\Omega = \begin{pmatrix} 0_n & I_n \\ -I_n & 0_n \end{pmatrix} \quad (16)$$

where  $I_n$  denotes the  $n \times n$  identity matrix and  $0_n$  denotes the  $n \times n$  zero matrix. Note that  $\Omega$  is antisymmetric and satisfies  $\Omega^2 = -I$ . We then define the **symplectic group**  $Sp(2n)$  by

$$Sp(2n) = \{A \in GL(2n): A^T \Omega A = \Omega\} . \quad (17)$$

The matrices  $A \in Sp(2n)$  are called **symplectic matrices**.<sup>3</sup>

Taking determinants of the defining equation  $A^T \Omega A = \Omega$  and using the nondegeneracy  $\det \Omega \neq 0$ ,<sup>4</sup> we conclude that every symplectic matrix satisfies

$$\det A = \pm 1 . \quad (18)$$

But, unlike in the case of the orthogonal group, it turns out that the case  $\det A = -1$  does *not* occur; all symplectic matrices in fact have  $\det A = +1$ . But the proof of this fact requires

---

<sup>2</sup>For a real antisymmetric matrix, the eigenvalues are pure imaginary and come in complex-conjugate pairs. If the dimension is odd, then at least one of the eigenvalues must be zero.

<sup>3</sup>In equation (39) of Handout #12 I used a slightly different convention, writing  $A \Omega A^T = \Omega$  instead of  $A^T \Omega A = \Omega$ .

<sup>4</sup>In fact  $\det \Omega = +1$ , but we don’t need this.

a deeper algebraic study, and I will not pursue it here.<sup>5</sup>

What is the dimension of the symplectic group  $Sp(2n)$ ? The defining equation  $A^T\Omega A = \Omega$  naively has  $(2n)^2 = 4n^2$  entries. But it is important to note that the matrix  $A^T\Omega A$  is *antisymmetric* (no matter what  $A$  is); therefore, some of the  $(2n)^2$  equations in  $A^T\Omega A = \Omega$  are redundant (namely, the  $ij$  equation says the same thing as the  $ji$  equation, and the diagonal equations say  $0 = 0$ ). So the non-redundant equations are, for instance, those above the diagonal (i.e.  $i < j$ ); there are  $(2n)(2n - 1)/2$  of these [you should check this carefully]. Therefore, the dimension of the symplectic group  $Sp(2n)$  is

$$(2n)^2 - \frac{(2n)(2n - 1)}{2} = \frac{(2n)(2n + 1)}{2} = n(2n + 1) . \quad (19)$$

## 2 Lie algebras

Let  $G$  be a Lie group of  $n \times n$  matrices: for instance, one of the “classical matrix groups” introduced in the preceding section. It turns out that much of the structure of  $G$  can be understood by looking only at the behavior of  $G$  in a tiny neighborhood of the identity element. Intuitively speaking, we consider matrices  $A = I + \epsilon\mathbf{A}$  with “ $\epsilon$  infinitesimal” and we ask under what conditions this matrix belongs to  $G$  “through first order in  $\epsilon$ ”. More rigorously, we consider a *smooth curve*  $A(t)$  lying in  $G$  and satisfying  $A(0) = I$ , and we ask what are the possible values of  $A'(0) = (d/dt)A(t)|_{t=0}$ .<sup>6</sup> Geometrically, what this means is that we regard  $G$  as a submanifold embedded in the vector space  $\mathbb{R}^{n \times n}$ , and we seek to determine the *tangent space* to  $G$  at the identity element  $I$ . The **Lie algebra**  $\mathfrak{g}$  associated to the Lie group  $G$  is, by definition, the tangent space to  $G$  at the point  $I$ ; that is, it is the set of all possible values of  $A'(0)$  when we consider all smooth curves  $A(t)$  lying in  $G$  and satisfying  $A(0) = I$ . Since a tangent space to a manifold is always a vector space (that is, it is closed under addition and multiplication by scalars), it follows that the Lie algebra  $\mathfrak{g}$  is a vector space. This fact makes it somewhat easier to study than  $G$  itself, which is a nonlinear manifold. That is the principal reason for introducing Lie algebras.

So our next task is to determine the Lie algebra  $\mathfrak{g}$  for each of the “classical matrix groups”  $G$  introduced in the preceding section.

### 2.1 The general linear group

The case of the general linear group  $GL(n)$  is easy: *every* matrix in a neighborhood of the identity matrix belongs to  $GL(n)$ , so the Lie algebra  $\mathfrak{gl}(n)$  consists of *all*  $n \times n$  matrices. That is,

$$\mathfrak{gl}(n) = \mathbb{R}^{n \times n} . \quad (20)$$

Of course  $\mathfrak{gl}(n)$  is a vector space of dimension  $n^2$ .

---

<sup>5</sup>The best way to prove this is to use the **pfaffian**, which is a kind of “square root of a determinant” for antisymmetric matrices. See e.g. [http://en.wikipedia.org/wiki/Symplectic\\_matrix](http://en.wikipedia.org/wiki/Symplectic_matrix) and <http://en.wikipedia.org/wiki/Pfaffian>

<sup>6</sup>Here  $t$  is simply a parameter; it need not be interpreted as “time”. But it *can* be interpreted as “time” if you wish: that is, we consider a “particle” moving through  $G$ , reaching  $I$  at time  $t = 0$ , and we ask what are its possible “velocities” at time  $t = 0$ .

## 2.2 The special linear group

The next case to consider is the special linear group  $SL(n)$ , which consists of the matrices  $A$  satisfying  $\det A = 1$ . Consider a smooth curve  $A(t)$  lying in  $SL(n)$  and satisfying  $A(0) = I$ , say

$$A(t) = I + \mathbf{A}t + \mathbf{A}_2 t^2 + \mathbf{A}_3 t^3 + \dots \quad (21)$$

As explained last week in class, we have

$$\det A(t) = 1 + (\operatorname{tr} \mathbf{A})t + O(t^2) \quad (22)$$

where  $\operatorname{tr} \mathbf{A} = \sum_{i=1}^n \mathbf{A}_{ii}$  is the **trace** of the matrix  $\mathbf{A}$ . [Recall the proof: we start from the definition of determinant as a sum over permutations  $\sigma$ . Every  $\sigma$  other than the identity permutation must contribute the product of *at least two* off-diagonal elements, so it will make a contribution starting only at order  $t^2$  or higher. The identity permutation will contribute the product of diagonal elements, which is  $\prod_{i=1}^n (1 + t\mathbf{A}_{ii}) = 1 + t \sum_{i=1}^n \mathbf{A}_{ii} + O(t^2)$ .] Therefore, if the curve  $A(t)$  lies in  $SL(n)$  — that is, if  $\det A(t) = 1$  for all  $t$  — then we must have  $\operatorname{tr} \mathbf{A} = 0$ . So every matrix in the Lie algebra  $\mathfrak{sl}(n)$  must be **traceless**. And conversely, if  $\mathbf{A}$  is a traceless  $n \times n$  matrix, then it can be shown that the particular curve

$$A(t) = e^{t\mathbf{A}} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{t^k \mathbf{A}^k}{k!} \quad (23)$$

satisfies  $\det A(t) = 1$  for all  $t$ . So the Lie algebra  $\mathfrak{sl}(n)$  consists precisely of the  $n \times n$  matrices that are traceless:

$$\mathfrak{sl}(n) = \{\mathbf{A} \in \mathbb{R}^{n \times n} : \operatorname{tr} \mathbf{A} = 0\} . \quad (24)$$

Note that  $\mathfrak{sl}(n)$  is a vector space of dimension  $n^2 - 1$  (why?).

**Remark.** The matrix exponential, defined by

$$e^M \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{M^k}{k!} , \quad (25)$$

has some but not all of the properties of the ordinary exponential; this is because matrix multiplication, unlike ordinary multiplication, is noncommutative. Thus, if  $M$  and  $N$  are two matrices, in general we do *not* have  $e^M e^N = e^{M+N}$ ; rather,  $\log(e^M e^N)$  is given by a much more complicated formula, called the **Campbell–Baker–Hausdorff formula**. But if  $M$  and  $N$  happen to *commute* (that is,  $MN = NM$ ), then it *is* true that  $e^M e^N = e^{M+N}$ . In particular, since all multiples of a single matrix  $\mathbf{A}$  trivially commute, it follows that  $e^{s\mathbf{A}} e^{t\mathbf{A}} = e^{(s+t)\mathbf{A}}$ , or in other words  $A(s)A(t) = A(s+t)$ . Thus, the curve  $A(t) = e^{\mathbf{A}t}$  forms a *commutative subgroup* of  $G$ : it is the **one-parameter subgroup** of  $G$  generated by the “infinitesimal generator”  $\mathbf{A}$ .

## 2.3 The orthogonal and special orthogonal groups

Now consider the orthogonal group  $O(n)$  and its subgroup  $SO(n)$ . Any smooth (or even continuous) curve in  $O(n)$  passing through the identity element must lie entirely in  $SO(n)$ , since the determinant obviously cannot jump from  $+1$  to  $-1$ . So  $O(n)$  and  $SO(n)$  will have the same Lie algebra, which we call  $\mathfrak{so}(n)$ .

So consider a smooth curve  $A(t)$  lying in  $O(n)$  and satisfying  $A(0) = I$ , say

$$A(t) = I + \mathbf{A}t + \mathbf{A}_2 t^2 + \mathbf{A}_3 t^3 + \dots \quad (26)$$

The condition to lie in  $O(n)$  is  $A^T A = I$ . Since

$$A(t)^T A(t) = I + (\mathbf{A} + \mathbf{A}^T)t + O(t^2), \quad (27)$$

we conclude that every matrix in the Lie algebra  $\mathfrak{so}(n)$  must satisfy  $\mathbf{A} + \mathbf{A}^T = 0$ , i.e. it must be **antisymmetric**. And conversely, if  $\mathbf{A}$  is a antisymmetric  $n \times n$  matrix, then it can be shown that

$$A(t) = e^{t\mathbf{A}} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{t^k \mathbf{A}^k}{k!} \quad (28)$$

satisfies  $\det A(t) = 1$  for all  $t$ . So the Lie algebra  $\mathfrak{so}(n)$  consists precisely of the  $n \times n$  matrices that are antisymmetric:

$$\mathfrak{so}(n) = \{ \mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A}^T = -\mathbf{A} \}. \quad (29)$$

Note that  $\mathfrak{so}(n)$  is a vector space of dimension  $n(n-1)/2$  [why?].

A similar procedure can be employed to determine the Lie algebra  $\mathfrak{so}(p, q)$  associated to the special pseudo-orthogonal group  $SO(p, q)$ . The details are left to you.

## 2.4 The symplectic group

Finally, we consider the symplectic group  $Sp(2n)$ , which you will recall is the set of  $2n \times 2n$  real matrices  $A$  satisfying  $A^T \Omega A = \Omega$ , where the antisymmetric matrix  $\Omega$  is defined by (16).

So consider a smooth curve  $A(t)$  lying in  $Sp(2n)$  and satisfying  $A(0) = I$ , say

$$A(t) = I + \mathbf{A}t + \mathbf{A}_2 t^2 + \mathbf{A}_3 t^3 + \dots \quad (30)$$

The condition to lie in  $Sp(2n)$  is  $A^T \Omega A = \Omega$ . Since

$$A(t)^T \Omega A(t) = I + (\mathbf{A}^T \Omega + \Omega \mathbf{A})t + O(t^2), \quad (31)$$

we conclude that every matrix in the Lie algebra  $\mathfrak{sp}(2n)$  must satisfy

$$\mathbf{A}^T \Omega + \Omega \mathbf{A} = 0. \quad (32)$$

This same equation arose already in Handout #12 when we were discussing “infinitesimal canonical transformations”.<sup>7</sup> Recall how we handled it: since  $\Omega$  is antisymmetric, we can also write the condition (32) as

$$\Omega \mathbf{A} - \mathbf{A}^T \Omega^T = 0 \quad (33)$$

or in other words

$$\Omega \mathbf{A} - (\Omega \mathbf{A})^T = 0. \quad (34)$$

So this says that the matrix  $\Omega \mathbf{A}$  is symmetric, or equivalently that the matrix

$$\Omega^T (\Omega \mathbf{A}) \Omega = \mathbf{A} \Omega \quad (35)$$

is symmetric. And conversely, if  $\mathbf{A}$  is a  $2n \times 2n$  matrix such that  $\Omega \mathbf{A}$  is symmetric, then it can be shown that

$$A(t) = e^{t\mathbf{A}} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{t^k \mathbf{A}^k}{k!} \quad (36)$$

satisfies  $A(t)^T \Omega A(t) = \Omega$  for all  $t$ . So the Lie algebra  $\mathfrak{sp}(2n)$  consists precisely of the  $2n \times 2n$  matrices such that  $\Omega \mathbf{A}$  is symmetric:

$$\mathfrak{sp}(2n) = \{ \mathbf{A} \in \mathbb{R}^{2n \times 2n} : \Omega \mathbf{A} = (\Omega \mathbf{A})^T \}. \quad (37)$$

Note that  $\mathfrak{sp}(2n)$  is a vector space of dimension

$$\frac{(2n)(2n+1)}{2} = n(2n+1) \quad (38)$$

(why?).

### 3 More on Lie algebras: The Lie bracket

The Lie algebra  $\mathfrak{g}$ , being the tangent space to the Lie group  $G$  at the identity element, is a vector space. But it is not *just* a vector space; it is also an **algebra**, i.e. a vector space equipped with a bilinear product. This bilinear product  $\mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$  is called the **Lie bracket**; it encodes much of the structure of the underlying Lie group  $G$ , and in a form that is easier to work with than the multiplication law of  $G$  itself.<sup>8</sup> The purpose of this section is to explain what the Lie bracket is and how it encodes the structure of  $G$ .

The most interesting groups are *noncommutative*: that is,  $gh \neq hg$  for at least some pairs  $g, h \in G$ . We can rewrite this as  $gh(hg)^{-1} \neq e$ , where  $e$  is the identity element of  $G$ . In group theory, the quantity  $gh(hg)^{-1} = ghg^{-1}h^{-1}$  is called the **commutator** of  $g$  and  $h$ ; its deviation from  $e$  measures in some sense the degree to which  $g$  fails to commute with  $h$ .

---

<sup>7</sup>With a transpose compared to the present notation: that is,  $K$  in Handout #12 [see equations (50) and following] corresponds to  $\mathbf{A}^T$  here. I have therefore slightly modified the treatment from Handout #12 in order to avoid unnecessary transposes in the final formula.

<sup>8</sup>More precisely, the Lie bracket encodes *all* of the *local* structure of the group  $G$ . But it obviously does not tell us about *global* aspects of  $G$ , such as the difference between the groups  $O(n)$  and  $SO(n)$ , both of which have the same Lie algebra.

Of course, in a general (e.g. discrete) group  $G$  there may not be any sensible way of defining quantitatively what we mean by the “deviation” of  $ghg^{-1}h^{-1}$  from  $e$ . But in a Lie group there is an obvious way of doing so, since the group elements are parametrized by a vector of real numbers. In particular, our examples of Lie groups are groups of  $n \times n$  real matrices (and  $e$  is the identity matrix  $I$ ), so we can simply look at the matrix difference  $ghg^{-1}h^{-1} - I$ : its deviation from zero tells us the degree to which  $g$  fails to commute with  $h$ .

Let us do this when  $g$  and  $h$  are elements of  $G$  near the identity element. More precisely, let us consider a smooth curve  $g(s)$  lying in  $G$  and satisfying  $g(0) = I$ , say

$$g(s) = I + \mathbf{A}s + O(s^2), \quad (39)$$

and another smooth curve  $h(t)$  lying in  $G$  and satisfying  $h(0) = I$ , say

$$h(t) = I + \mathbf{B}t + O(t^2). \quad (40)$$

We now wish to compute the commutator  $g(s)h(t)g(s)^{-1}h(t)^{-1}$  in power series in  $s$  and  $t$ .

Recall first that for any real (or complex) number  $x$  satisfying  $|x| < 1$ , we have

$$(1 + x)^{-1} = 1 - x + x^2 - x^3 + x^4 - \dots \quad (41)$$

in the sense that the power series on the right-hand side is absolutely convergent and equals  $(1 + x)^{-1}$ . [The absolute convergence is easy to prove, using properties of geometric series. To show that the infinite series equals  $(1 + x)^{-1}$ , just multiply it by  $1 + x$ , rearrange the terms (this is legitimate since the series is absolutely convergent), and see that the result is  $1 + 0x + 0x^2 + \dots$ .] Similarly, for any real (or complex)  $n \times n$  matrix  $A$  satisfying  $\|A\| < 1$  (where  $\|\cdot\|$  is a suitable norm), we have

$$(I + A)^{-1} = I - A + A^2 - A^3 + A^4 - \dots; \quad (42)$$

the proof is exactly the same. This sum is called the **Neumann series** for the matrix inverse. (Here we won't bother to make explicit what norm  $\|\cdot\|$  we are using; but it won't matter, because anyway we will be using the Neumann formula only for matrices  $A$  that are *very* small.)

So using the Neumann formula we have

$$g(s)^{-1} = I - \mathbf{A}s + O(s^2) \quad (43)$$

and

$$h(t)^{-1} = I - \mathbf{B}t + O(t^2) \quad (44)$$

[why?]. The commutator of  $g(s)$  and  $h(t)$  is therefore

$$\begin{aligned} g(s)h(t)g(s)^{-1}h(t)^{-1} &= [I + \mathbf{A}s + O(s^2)][I + \mathbf{B}t + O(t^2)][I - \mathbf{A}s + O(s^2)][I - \mathbf{B}t + O(t^2)] \\ &\quad (45a) \end{aligned}$$

$$\begin{aligned} &= I + (\mathbf{A}\mathbf{B} - \mathbf{B}\mathbf{A})st + \text{terms of order } st^2 \text{ and } s^2t \text{ and higher.} \\ &\quad (45b) \end{aligned}$$

(You should check this computation carefully!) Note that there are no terms  $s^n t^0$  with  $n \geq 1$ , because  $g(s)h(0)g(s)^{-1}h(0)^{-1} = I$  for all  $s$  [why?]; and likewise there are no terms  $s^0 t^n$  with  $n \geq 1$ . So all the terms not explicitly shown here must be of order  $st^2$  or  $s^2 t$  or higher.

From (45) we see that the quantity  $\mathbf{AB} - \mathbf{BA}$  thus measures the degree to which  $g(s)$  fails to commute with  $h(t)$ , at the lowest nontrivial order (which is order  $st$ ). We therefore make the definition: the **Lie bracket**  $[\mathbf{A}, \mathbf{B}]$  is defined by<sup>9</sup>

$$[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA} . \quad (46)$$

The Lie bracket measures the noncommutativity of the Lie group  $G$  in a small neighborhood of the identity element.

Since the curves  $g(s)$  and  $h(t)$  lie in the group  $G$ , their “infinitesimal elements”  $\mathbf{A}$  and  $\mathbf{B}$  must belong to the Lie algebra  $\mathfrak{g}$  (by definition); and conversely, any two elements  $\mathbf{A}$  and  $\mathbf{B}$  of the Lie algebra  $\mathfrak{g}$  are tangent vectors to suitable smooth curves  $g(s)$  and  $h(t)$  lying in  $G$ . What about  $[\mathbf{A}, \mathbf{B}]$ ? We see from (45) that  $[\mathbf{A}, \mathbf{B}]$  is indeed the tangent vector to a curve lying in  $G$ , namely the curve

$$g(\sqrt{t})h(\sqrt{t})g(\sqrt{t})^{-1}h(\sqrt{t})^{-1} = I + [\mathbf{A}, \mathbf{B}]t + O(t^{3/2}) . \quad (47)$$

(I am not sure whether the  $O(t^{3/2})$  here can actually be replaced by  $O(t^2)$ ; but anyway, formula (47) already suffices to show that the curve in question is at least *once* continuously differentiable, with tangent vector  $[\mathbf{A}, \mathbf{B}]$  at  $t = 0$ .) It follows that  $[\mathbf{A}, \mathbf{B}]$  belongs to the Lie algebra  $\mathfrak{g}$ .

**Conclusion:** The Lie bracket  $[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$  maps  $\mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ .

The Lie bracket has the following fundamental properties:

1. **Bilinearity.** We have

$$[\alpha_1 \mathbf{A}_1 + \alpha_2 \mathbf{A}_2, \mathbf{B}] = \alpha_1 [\mathbf{A}_1, \mathbf{B}] + \alpha_2 [\mathbf{A}_2, \mathbf{B}] \quad (48)$$

and likewise for  $\mathbf{B}$ .

2. **Anticommutativity.** We have

$$[\mathbf{A}, \mathbf{B}] = -[\mathbf{B}, \mathbf{A}] . \quad (49)$$

In particular it follows that  $[\mathbf{A}, \mathbf{A}] = 0$  (why?).

3. **Jacobi identity.** We have

$$[\mathbf{A}, [\mathbf{B}, \mathbf{C}]] + [\mathbf{B}, [\mathbf{C}, \mathbf{A}]] + [\mathbf{C}, [\mathbf{A}, \mathbf{B}]] = 0 \quad (50)$$

or equivalently (using anticommutativity)

$$[[\mathbf{A}, \mathbf{B}], \mathbf{C}] + [[\mathbf{B}, \mathbf{C}], \mathbf{A}] + [[\mathbf{C}, \mathbf{A}], \mathbf{B}] = 0 \quad (51)$$

---

<sup>9</sup>The Lie bracket is also called the **commutator** of the two matrices  $\mathbf{A}$  and  $\mathbf{B}$ . It is unfortunate that the same word “commutator” is used in group theory and in linear algebra for two closely related but slightly different notions.

The bilinearity and anticommutativity are trivial, and the Jacobi identity is an easy computation (do it!).

This structure is so important in mathematics that it is given a name: any vector space  $V$  equipped with a product  $V \times V \rightarrow V$  that is bilinear, anticommutative and satisfies the Jacobi identity is called a **Lie algebra** (or an **abstract Lie algebra** if we wish to distinguish it from the concrete Lie algebras of  $n \times n$  matrices introduced here).

## 4 The two-dimensional rotation group $SO(2)$

Let  $R(\theta)$  be the rotation (about the origin) by angle  $\theta$  in the plane  $\mathbb{R}^2$ , where a positive angle is taken as corresponding to an anti-clockwise rotation. Simple trigonometry shows that the action on the unit vectors  $\hat{\mathbf{e}}_x$  and  $\hat{\mathbf{e}}_y$  is

$$R(\theta)\hat{\mathbf{e}}_x = (\cos\theta)\hat{\mathbf{e}}_x + (\sin\theta)\hat{\mathbf{e}}_y \quad (52a)$$

$$R(\theta)\hat{\mathbf{e}}_y = -(\sin\theta)\hat{\mathbf{e}}_x + (\cos\theta)\hat{\mathbf{e}}_y \quad (52b)$$

(draw the picture for yourself!). The corresponding  $2 \times 2$  matrix is therefore

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \quad (53)$$

The rotation group  $SO(2)$  consists of the matrices  $R(\theta)$  for  $\theta \in \mathbb{R}$ . Of course, we have  $R(\theta) = R(\theta + 2\pi n)$  for all integers  $n$ , so it suffices to consider (for instance)  $-\pi < \theta \leq \pi$ .

The group law

$$R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2) \quad (54)$$

is an easy consequence of trigonometric angle-addition formulae (check it!). In particular,  $R(\theta_1)R(\theta_2) = R(\theta_2)R(\theta_1)$ : that is, any two rotations (about the origin) in the plane commute. The rotation group  $SO(2)$  is therefore *abelian*.

The Lie algebra  $\mathfrak{so}(2)$  consists of  $2 \times 2$  antisymmetric matrices. These are multiples of the generator

$$\mathbf{L} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \left. \frac{d}{d\theta} R(\theta) \right|_{\theta=0}. \quad (55)$$

Of course,  $[\mathbf{L}, \mathbf{L}] = 0$ , reflecting the fact that  $SO(2)$  is abelian.

This trivial behavior of the rotation group  $SO(2)$  is, of course, special to two dimensions. For  $n \geq 3$ , the rotation group  $SO(n)$  is nonabelian, as we shall now see.

## 5 The three-dimensional rotation group $SO(3)$

The rotation group  $SO(3)$  consists of  $3 \times 3$  real matrices that are orthogonal ( $A^T A = I$ ) and have determinant  $+1$ . Simple examples of such matrices are given by rotations about the coordinate axes. For instance, a rotation of angle  $\theta$  around the  $z$  axis (i.e. in the  $x, y$ -plane)

acts on the unit vectors  $\hat{\mathbf{e}}_x$ ,  $\hat{\mathbf{e}}_y$ ,  $\hat{\mathbf{e}}_z$  by

$$R_z(\theta) \hat{\mathbf{e}}_x = (\cos \theta) \hat{\mathbf{e}}_x + (\sin \theta) \hat{\mathbf{e}}_y \quad (56a)$$

$$R_z(\theta) \hat{\mathbf{e}}_y = -(\sin \theta) \hat{\mathbf{e}}_x + (\cos \theta) \hat{\mathbf{e}}_y \quad (56b)$$

$$R_z(\theta) \hat{\mathbf{e}}_z = \hat{\mathbf{e}}_z \quad (56c)$$

These become the *columns* of the matrix  $R_z(\theta)$ , so we have

$$R_z(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (57)$$

(We always consider our rotation matrices  $R$  as acting on *column vectors*  $\mathbf{v}$ , by the action  $\mathbf{v} \mapsto R\mathbf{v}$ .) This, of course, is nothing other than the rotation (52)/(53) acting in the  $x, y$ -plane, together with an identity action on the  $z$  coordinate. Please note that, according to the right-hand rule, a rotation whose axis is in the  $+z$  direction corresponds to a rotation of the  $+x$  direction towards the  $+y$  direction, i.e. a rotation of positive angle as we defined it in the previous section.

Similarly, a rotation of angle  $\theta$  around the  $+x$  axis (i.e. in the  $y, z$ -plane) is given by the matrix

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix}. \quad (58)$$

Here, according to the right-hand rule, a rotation whose axis is in the  $+x$  direction corresponds to a rotation of the  $+y$  direction towards the  $+z$  direction.

Finally, a rotation of angle  $\theta$  around the  $+y$  axis (i.e. in the  $x, z$ -plane) is given by the matrix

$$R_y(\theta) = \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix}. \quad (59)$$

Note that here, according to the right-hand rule, a rotation whose axis is in the  $+y$  direction corresponds to a rotation of the  $+z$  direction towards the  $+x$  direction — you should check carefully the signs in this formula to make sure you understand them!

It turns out that the group  $SO(3)$  is *generated* by these special matrices: that is, every element of  $SO(3)$  is a finite product of these. Indeed, it turns out that every element of  $SO(3)$  can be written as a product of *three* coordinate-axis rotation matrices, for instance as

$$A = R_z(\psi) R_x(\theta) R_z(\varphi). \quad (60)$$

The angles  $\varphi, \theta, \psi$  are called **Euler angles**. There are, in fact, many different conventions for Euler angles, and one has to check carefully which convention any given author is using. More details can be found, for instance, at [https://en.wikipedia.org/wiki/Euler\\_angles](https://en.wikipedia.org/wiki/Euler_angles)

It is easy to check that the rotation group  $SO(3)$  is *nonabelian*: rotations about two different axes do *not* in general commute. For instance, you can check that

$$R_x(\theta_1) R_z(\theta_2) \neq R_z(\theta_2) R_x(\theta_1) \quad (61)$$

whenever  $\theta_1$  and  $\theta_2$  are not multiples of  $\pi$ . Indeed,  $R_x(\pi/2)$  does not commute with  $R_z(\pi/2)$  — as I demonstrated in class by carrying out  $90^\circ$  rotations of an eraser.

The Lie algebra  $\mathfrak{so}(3)$  consists of  $3 \times 3$  antisymmetric matrices. This is a 3-dimensional vector space, and a basis for this vector space is given by the three generators

$$\mathbf{L}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = \left. \frac{d}{d\theta} R_x(\theta) \right|_{\theta=0} \quad (62a)$$

$$\mathbf{L}_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} = \left. \frac{d}{d\theta} R_y(\theta) \right|_{\theta=0} \quad (62b)$$

$$\mathbf{L}_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \left. \frac{d}{d\theta} R_z(\theta) \right|_{\theta=0} \quad (62c)$$

These matrices have the Lie brackets

$$[\mathbf{L}_x, \mathbf{L}_y] = \mathbf{L}_z \quad (63a)$$

$$[\mathbf{L}_y, \mathbf{L}_z] = \mathbf{L}_x \quad (63b)$$

$$[\mathbf{L}_z, \mathbf{L}_x] = \mathbf{L}_y \quad (63c)$$

and their reversals by anticommutativity (e.g.  $[\mathbf{L}_y, \mathbf{L}_x] = -\mathbf{L}_z$ ), along with the trivial commutators  $[\mathbf{L}_x, \mathbf{L}_x] = [\mathbf{L}_y, \mathbf{L}_y] = [\mathbf{L}_z, \mathbf{L}_z] = 0$ . You should check these commutators, just to make sure that we have gotten all the signs right!

## More details concerning Euler angles

As I said, there are a variety of conventions concerning Euler angles, and one has to check carefully which convention any given author is using.

Things are made even more confusing by the fact that there are two different ways of interpreting rotations: as **active transformations** or as **passive transformations**.<sup>10</sup> In both approaches, we fix a coordinate system  $x, y, z$  attached to the earth and another coordinate system  $X, Y, Z$  attached to the rigid body; but the two approaches differ in how we interpret the rotation matrix. In the active point of view, we interpret the transformation as specifying the physical orientation of the rigid body, pivoted at the origin, relative to the earth axes: that is,  $\hat{\mathbf{e}}_X = R\hat{\mathbf{e}}_x$ ,  $\hat{\mathbf{e}}_Y = R\hat{\mathbf{e}}_y$ ,  $\hat{\mathbf{e}}_Z = R\hat{\mathbf{e}}_z$ . In the passive point of view, by contrast, we consider some fixed point  $P$  in space, and we ask what coordinates  $(x, y, z)$  the earth axes would attribute to that point, and what coordinates  $(X, Y, Z)$  the rigid-body axes

would attribute to that same point; the passive transformation tells us how to go from  $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$

---

<sup>10</sup>See [https://en.wikipedia.org/wiki/Active\\_and\\_passive\\_transformation](https://en.wikipedia.org/wiki/Active_and_passive_transformation)

to  $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$ . To see how the two approaches are connected, note first that

$$P = x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y + z\hat{\mathbf{e}}_z = X\hat{\mathbf{e}}_X + Y\hat{\mathbf{e}}_Y + Z\hat{\mathbf{e}}_Z. \quad (64)$$

We can rewrite the two sides of this equation as column vectors with respect to the basis  $\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y, \hat{\mathbf{e}}_z$ :

$$x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y + z\hat{\mathbf{e}}_z = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (65a)$$

$$X\hat{\mathbf{e}}_X + Y\hat{\mathbf{e}}_Y + Z\hat{\mathbf{e}}_Z = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (65b)$$

(Make sure that you understand this last equation: the *columns* of the matrix  $R$  are the vectors  $\hat{\mathbf{e}}_X, \hat{\mathbf{e}}_Y, \hat{\mathbf{e}}_Z$  written with respect to the basis  $\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y, \hat{\mathbf{e}}_z$ ; when we apply this matrix to

the column vector  $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$ , we obtain  $X$  times the first column plus  $Y$  times the second column plus  $Z$  times the third column.) Putting these two equations together, we obtain

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \quad (66)$$

or equivalently

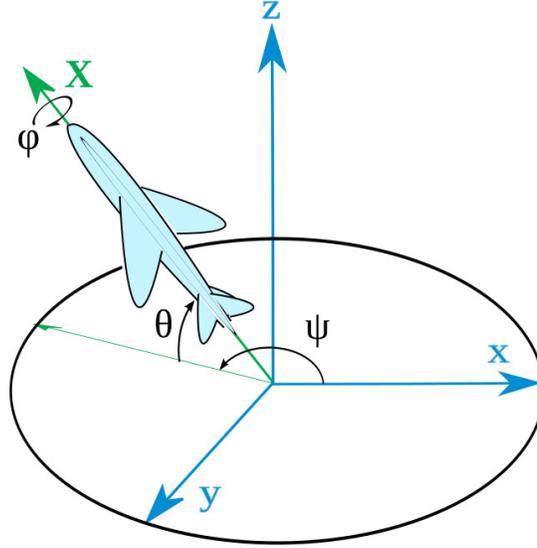
$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \quad (67)$$

That is, in any given physical situation, the matrix associated to the passive interpretation of the transformation is the *inverse* of the matrix associated to the active interpretation of the transformation.

In what follows, **I shall always use the active interpretation of the transformation  $R$ .** (But be warned: many physics books use the passive interpretation!)

Let us now consider one convention for the Euler angles; though it's not the one that is most commonly used in physics, it's one that's easy to visualize and interpret.

Imagine that the rigid body is an airplane. Attached to the earth we have axes  $x$  (east),  $y$  (north) and  $z$  (up). Attached to the airplane we have axes  $X$  (forwards),  $Y$  (left) and  $Z$  (up). (Note that both of these are indeed right-handed systems, as required.) The airplane is originally parked so that the  $X, Y, Z$  axes coincide with the  $x, y, z$  axes: that is, it is horizontal and facing east. Now we make three rotations:



Taken from [https://en.wikipedia.org/wiki/Euler\\_angles](https://en.wikipedia.org/wiki/Euler_angles)

**Warning:** What is shown in this picture as the  $y$  axis is really the  $-y$  axis, according to the right-hand rule.

First, we rotate by an angle  $\psi$  in the  $x$ - $y$  plane, where a positive angle corresponds to a rotation of the  $+x$  direction towards the  $+y$  direction. In other words, this is a rotation about the  $+z$  axis. This rotation is implemented by the rotation matrix

$$R_1 = R_z(\psi) . \quad (68)$$

The airplane is now pointing in the direction  $\psi$  (in the sense of plane polar coordinates), still horizontal. The axes in the airplane are now pointing in directions that we shall call  $x', y', z'$ ; of course  $z'$  is the same as  $z$ .

Next we make the plane's nose point upwards by an angle  $\theta$ . (Or rather, upwards if  $\theta > 0$ , downwards if  $\theta < 0$ .) That is, we rotate by an angle  $\theta$  in the  $x'$ - $z'$  plane, where a positive angle corresponds to a rotation of the  $+x'$  direction towards the  $+z'$  direction. In other words, this is a rotation about the  $-y'$  axis. (Make sure you understand the sign here, by using the right-hand rule!) Equivalently, we can say that this is a rotation by an angle  $-\theta$  around the  $+y'$  axis.

Now here is the trick: a rotation about the  $+y'$  axis can be implemented by first taking the airplane back to its original position (using the map  $R_1^{-1}$ ), then making the rotation about the  $+y$  axis, then undoing what we did by using the map  $R_1$ . So we have

$$R_2 = R_1 R_y(-\theta) R_1^{-1} \quad (69)$$

(notice that the three operations occur in order *from right to left*). The axes in the airplane are now pointing in directions that we shall call  $x'', y'', z''$ .

Finally, we rotate the plane by an angle  $\varphi$  around its front-to-back axis, where  $\varphi > 0$  corresponds (as in the Figure) to rotating the left side of the plane downwards, i.e. rotating the  $+y''$  direction towards the  $-z''$  direction. In other words, this is a rotation about the  $-x''$  axis. Equivalently, we can say that this is a rotation by an angle  $-\varphi$  around the  $+x''$  axis.

Now we apply the same trick: a rotation about the  $+x''$  axis can be implemented by first taking the airplane back to its original position (using the map  $(R_2R_1)^{-1} = R_1^{-1}R_2^{-1}$ ), then making the rotation about the  $+x$  axis, then undoing what we did by using the map  $R_2R_1$ . So we have

$$R_3 = R_2R_1 R_x(-\varphi) (R_2R_1)^{-1} \quad (70)$$

(again the three operations occur in order *from right to left*). The axes in the airplane are now pointing in directions  $X, Y, Z$ .

The transformation that gets us from the initial to the final position is

$$R_3R_2R_1 = R_2R_1 R_x(-\varphi) = R_1 R_y(-\theta) R_x(-\varphi) = R_z(\psi) R_y(-\theta) R_x(-\varphi). \quad (71)$$

Note the order of the three factors.

If we had chosen to do what the Wikipedia article calls **extrinsic rotations** — that is, rotations about the coordinate axes  $x, y, z$  attached to the earth (in this case  $z$ , then  $y$ , then  $x$ ) — then we would have had these three rotation matrices in order *from right to left* (i.e., the usual order of writing matrices associated to successive linear transformations acting on a column vector).

But we chose instead to do **intrinsic rotations**, that is, rotations about the (temporary) coordinate axes attached to the body (in this case  $z$ , then  $y'$ , then  $x''$ ), since their geometric interpretation seems more natural. Then it turns out that we simply have these same matrices in the opposite order, i.e. *from left to right*.

Finally, we need to specify the *range* of each of the three angles. The angles  $\psi$  and  $\varphi$  are specified modulo  $2\pi$ , so they can belong to any interval of length  $2\pi$  one chooses: e.g. either  $[0, 2\pi]$  or  $[-\pi, \pi]$ . The angle  $\theta$  is like the colatitude coordinate in spherical polar coordinates: it ranges over an interval of length  $\pi$ , where here  $[-\pi/2, \pi/2]$  looks most appropriate to our geometrical interpretation (but  $[0, \pi]$  is also acceptable). It is *not* to be considered modulo  $\pi$ .