A Concrete Introduction to Classical Lie Groups Via the Exponential Map

Jean Gallier
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
e-mail: jean@saul.cis.upenn.edu

February 20, 2002

1 Introduction

The purpose of these notes is to give a concrete introduction to Lie groups and Lie algebras. Our ulterior motive is to present some beautiful mathematical concepts that can also be used as tools for solving practical problems arising in computer science, more specifically in robotics, motion planning, computer vision, computer graphics. Most texts on Lie groups and Lie algebras begin with prerequesites in differential geometry that are often formidable to average computer scientists (or average scientists, whatever that means!). We have also banged our head against the wall for a long time, trying to figure out what Lie groups and Lie algebras are all about, but recently, we realized that there is perhaps a way to break down the obstacles. We claim that one can sneak into the wonderful world of Lie groups and Lie algebras by playing with explicit matrix groups such as the group of rotations in $\mathbb{R}^2$ (or $\mathbb{R}^3$), and with the exponential map. After actually computing the exponential $A = e^B$ of a 2×2 skew symmetric matrix B and observing that it is a rotation matrix, and similarly for a 3×3 skew symmetric matrix B, one begins to suspect that there is something deep going on. Similarly, after the discovery that every real invertible $n \times n$ matrix A can be written as $A = RP$, where R is an orthogonal matrix and P is a positive definite symmetric matrix, and that P can be written as $P = e^S$ for some symmetric matrix S, one begins to appreciate the exponential map.

We attempt to give an elementary and concrete introduction to Lie groups and Lie algebras by studying a number of the so-called *classical groups*, such as the general linear group $\mathbf{GL}(n, \mathbb{R})$, the special linear group $\mathbf{SL}(n, \mathbb{R})$, the orthogonal group $\mathbf{O}(n)$, the special orthogonal group $\mathbf{SO}(n)$, and the group of affine rigid motions $\mathbf{SE}(n)$, and their Lie algebras $\mathfrak{gl}(n, \mathbb{R})$ (all matrices), $\mathfrak{sl}(n, \mathbb{R})$ (matrices with null trace), $\mathfrak{o}(n)$, and $\mathfrak{so}(n)$ (skew symmetric matrices). We also consider the corresponding groups of complex matrices and their Lie algebras. Whenever possible, we show that the exponential map is surjective. For this, all we need is some results of linear algebra about various normal forms for symmetric matrices and skew symmetric matrices. Thus, we begin by proving that there are nice normal forms (block diagonal matrices where the blocks have size at most two) for normal matrices and other special cases (symmetric matrices, skew symmetric matrices, orthogonal matrices). We also prove the spectral theorem for complex normal matrices. Having done that, we have all the tools to present the important *singular value decomposition* (SVD) and the *polar form* of a matrix, and we can't resist the temptation to explain these neat results. Then, we move on to the exponential map and show how various kinds of matrices are obtained as exponentials of others, leading the way to Lie groups and Lie algebras. On the way, we derive the classical "Rodrigues-like" formulae for rotations and for rigid motions in $\mathbb{R}^2$ and $\mathbb{R}^3$. We give an elementary proof that the exponential map is surjective for both $\mathbf{SO}(n)$ and $\mathbf{SE}(n)$, not using any topology, just our normal forms for matrices. The last section gives a quick introduction to Lie groups and Lie algebras. We define manifolds as embedded submanifolds of $\mathbb{R}^N$, and we only define linear Lie groups, using the famous result of Cartan (apparently actually due to Von Neumann) that a closed subgroup of $\mathbf{GL}(n, \mathbb{R})$ is a manifold, and thus a Lie group. This way, Lie algebras can be "computed" using tangent vectors to curves of the form $t \mapsto A(t)$, where $A(t)$ is a matrix. We do not pretend that any of this material is original (except perhaps for our proof of surjectivity of the exponential), but we believe that out point of view is somewhat original. In any case, we hope that these notes will be useful to our readers, and will inspire them to learn more about these beautiful (and useful!) theories. We intend to write more about applications of Lie groups and Lie algebras to computer science problems (in particular, motion interpolation) in a sequel of these notes.

2 Normal Linear Maps

First, let us recall the definition of an inner product on a real vector space. This section and the next two were inspired by Lang [15], Artin [2], Mac Lane and Birkhoff [16], Berger [3], and Bertin [5].

Definition 2.1 A real vector space E is a *Euclidean space* if it is equipped with a symmetric bilinear form $\varphi\colon E \times E \to \mathbb{R}$, which is also *positive definite*,[1] which means that

$$\varphi(\overrightarrow{u}, \overrightarrow{u}) > 0, \quad \text{for every } \overrightarrow{u} \neq \overrightarrow{0}.$$

More explicitly, $\varphi\colon E \times E \to \mathbb{R}$ satisfies the following axioms:

$$\begin{aligned}
\varphi(\overrightarrow{u_1} + \overrightarrow{u_2}, \overrightarrow{v}) &= \varphi(\overrightarrow{u_1}, \overrightarrow{v}) + \varphi(\overrightarrow{u_2}, \overrightarrow{v}),\\
\varphi(\overrightarrow{u}, \overrightarrow{v_1} + \overrightarrow{v_2}) &= \varphi(\overrightarrow{u}, \overrightarrow{v_1}) + \varphi(\overrightarrow{u}, \overrightarrow{v_2}),\\
\varphi(\lambda\overrightarrow{u}, \overrightarrow{v}) &= \lambda\varphi(\overrightarrow{u}, \overrightarrow{v}),\\
\varphi(\overrightarrow{u}, \lambda\overrightarrow{v}) &= \lambda\varphi(\overrightarrow{u}, \overrightarrow{v}),\\
\varphi(\overrightarrow{u}, \overrightarrow{v}) &= \varphi(\overrightarrow{v}, \overrightarrow{u}),\\
\overrightarrow{u} \neq \overrightarrow{0} \text{ implies that } \ &\varphi(\overrightarrow{u}, \overrightarrow{u}) > 0.
\end{aligned}$$

The real number $\varphi(\overrightarrow{u}, \overrightarrow{v})$ is also called the *inner product (or scalar product) of* $\overrightarrow{u}$ *and* $\overrightarrow{v}$. We also define the *quadratic form associated with* φ as the function $\Phi\colon E \to \mathbb{R}_+$ such that

$$\Phi(\overrightarrow{u}) = \varphi(\overrightarrow{u}, \overrightarrow{u}),$$

for all $\overrightarrow{u} \in E$.

Since φ is bilinear, we have $\varphi(\overrightarrow{0}, \overrightarrow{0}) = 0$, and since it is positive definite, we have the stronger fact that

$$\varphi(\overrightarrow{u}, \overrightarrow{u}) = 0 \quad \text{iff} \quad \overrightarrow{u} = \overrightarrow{0},$$

[1] A bilinear form $\varphi\colon E \times E \to \mathbb{R}$ is *definite* if for every $\overrightarrow{u} \in E$, $\overrightarrow{u} \neq \overrightarrow{0}$ implies that $\varphi(\overrightarrow{u}, \overrightarrow{u}) \neq 0$, and positive if for every $\overrightarrow{u} \in E$, $\varphi(\overrightarrow{u}, \overrightarrow{u}) \geq 0$.

that is $\Phi(\overrightarrow{u}) = 0$ iff $\overrightarrow{u} = \overrightarrow{0}$.

The standard example of a Euclidean space is $\mathbb{R}^n$, under the inner product φ defined such that

$$\varphi((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

An inner product on a Euclidean space is often denoted as $\langle -, - \rangle$. Also, recall that every linear map $f \colon E \to E$ has an *adjoint* f^* which is a linear map $f^* \colon E \to E$ such that

$$\langle f(\overrightarrow{u}), \overrightarrow{v} \rangle = \langle \overrightarrow{u}, f^*(\overrightarrow{v}) \rangle,$$

for all $\overrightarrow{u}, \overrightarrow{v} \in E$. Since $\langle -, - \rangle$ is symmetric, it is obvious that $f^{**} = f$. Given two Euclidean spaces E and F, where the inner product on E is denoted as $\langle -, - \rangle_1$ and the inner product on F is denoted as $\langle -, - \rangle_2$, given any linear map $f \colon E \to F$, there is a unique linear map $f^* \colon F \to E$ such that

$$\langle f(\overrightarrow{u}), \overrightarrow{v} \rangle_2 = \langle \overrightarrow{u}, f^*(\overrightarrow{v}) \rangle_1$$

for all $\overrightarrow{u} \in E$ and all $\overrightarrow{v} \in F$. The linear map f^* is also called the adjoint of f. This more general situation will be encountered when we deal with the singular value decomposition of rectangular matrices.

Definition 2.2 Given a Euclidean space E, a linear map $f \colon E \to E$ is *normal* if

$$f \circ f^* = f^* \circ f.$$

A linear map $f \colon E \to E$ is *self-adjoint* if $f = f^*$, *skew self-adjoint* if $f = -f^*$, and *orthogonal* if $f \circ f^* = f^* \circ f = \mathrm{id}$.

Obviously, a self-adjoint, skew self-adjoint, or orthogonal, linear map is a normal linear map. Our first goal is to show that for every normal linear map $f \colon E \to E$, there is an orthonormal basis (w.r.t. $\langle -, - \rangle$) such that the matrix of f over this basis has an especially nice form: it is a block diagonal matrix in which the blocks are either one-dimensional matrices (i.e., single entries) or two-dimensional matrices of the form

$$\begin{pmatrix} \lambda & \mu \\ -\mu & \lambda \end{pmatrix}$$

This normal form can be further refined if f is self-adjoint, skew self-adjoint, or orthogonal. As a first step, we show that f and f^* have the same kernel when f is normal.

Lemma 2.3 *Given a Euclidean space E, if $f \colon E \to E$ is a normal linear map, then* $\mathrm{Ker}\, f = \mathrm{Ker}\, f^*$.

Proof. First, let us prove that

$$\langle f(\vec{u}), f(\vec{v})\rangle = \langle f^*(\vec{u}), f^*(\vec{v})\rangle$$

for all $\vec{u}, \vec{v} \in E$. Since f^* is the adjoint of f and $f \circ f^* = f^* \circ f$, we have

$$\begin{aligned} \langle f(\vec{u}), f(\vec{u})\rangle &= \langle \vec{u}, (f^* \circ f)(\vec{u})\rangle, \\ &= \langle \vec{u}, (f \circ f^*)(\vec{u})\rangle, \\ &= \langle f^*(\vec{u}), f^*(\vec{u})\rangle. \end{aligned}$$

Since $\langle -, -\rangle$ is positive definite,

$$\langle f(\vec{u}), f(\vec{u})\rangle = 0 \quad \text{iff} \quad f(\vec{u}) = \vec{0},$$

$$\langle f^*(\vec{u}), f^*(\vec{u})\rangle = 0 \quad \text{iff} \quad f^*(\vec{u}) = \vec{0},$$

and since

$$\langle f(\vec{u}), f(\vec{u})\rangle = \langle f^*(\vec{u}), f^*(\vec{u})\rangle,$$

we have

$$f(\vec{u}) = \vec{0} \quad \text{iff} \quad f^*(\vec{u}) = \vec{0}.$$

Consequently, $\mathrm{Ker}\, f = \mathrm{Ker}\, f^*$. □

The next step is to show that for every linear map $f\colon E \to E$, there is some subspace W of dimension 1 or 2 such that $f(W) \subseteq W$. When $\dim(W) = 1$, the subspace W is actually an eigenspace for some real eigenvalue of f. Furthermore, when f is normal, there is a subspace W of dimension 1 or 2 such that $f(W) \subseteq W$ and $f^*(W) \subseteq W$. The difficulty is that the eigenvalues of f are not necessarily real. One way to get around this problem is to complexify both the vector space E and the inner product $\langle -, -\rangle$.

First, we need to embed a real vector space E into a complex vector space $E_{\mathbb{C}}$. A quick but somewhat bewildering way to do so is to define the complexification of E as the tensor product $\mathbb{C} \otimes E$. A more tangible way is to define the following structure.

Definition 2.4 Given a real vector space E, let $E_{\mathbb{C}}$ be the structure $E \times E$ under the addition operation

$$(\vec{u_1}, \vec{u_2}) + (\vec{v_1}, \vec{v_2}) = (\vec{u_1} + \vec{v_1}, \vec{u_2} + \vec{v_2}),$$

and multiplication by a complex scalar $z = x + iy$ defined such that

$$(x + iy) \cdot (\vec{u}, \vec{v}) = (x\vec{u} - y\vec{v}, y\vec{u} + x\vec{v}).$$

It is easily shown that the structure $E_{\mathbb{C}}$ is a complex vector space. It is also immediate that

$$(\vec{0}, \vec{v}) = i(\vec{v}, \vec{0}),$$

and thus, identifying E with the subspace of $E_{\mathbb{C}}$ consisting of all vectors of the form $(\overrightarrow{u}, \overrightarrow{0})$, we can write

$$(\overrightarrow{u}, \overrightarrow{v}) = \overrightarrow{u} + i\overrightarrow{v}.$$

Given a vector $\overrightarrow{w} = \overrightarrow{u} + i\overrightarrow{v}$, its *conjugate* $\overline{\overrightarrow{w}}$ is the vector $\overline{\overrightarrow{w}} = \overrightarrow{u} - i\overrightarrow{v}$. Then, conjugation is a map from $E_{\mathbb{C}}$ to itself which is an involution. If $(\overrightarrow{e}_1, \ldots, \overrightarrow{e}_n)$ is any basis of E, then $((\overrightarrow{e}_1, \overrightarrow{0}), \ldots, (\overrightarrow{e}_n, \overrightarrow{0}))$ is a basis of $E_{\mathbb{C}}$. We call such a basis a *real basis*.

Given a linear map $f\colon E \to E$, the map f can be extended to a linear map $f_{\mathbb{C}}\colon E_{\mathbb{C}} \to E_{\mathbb{C}}$ defined such that

$$f_{\mathbb{C}}(\overrightarrow{u} + i\overrightarrow{v}) = f(\overrightarrow{u}) + if(\overrightarrow{v}).$$

Next, we need to extend the inner product on E to an inner product on $E_{\mathbb{C}}$. Let us recall the definition of a Hermitian space.

Definition 2.5 A complex vector space E is a *Hermitian space* if it is equipped with a map $\varphi\colon E \times E \to \mathbb{C}$ which is linear in its first argument and semi-linear in its second argument (sometimes called a *sesquilinear map*), and which is also *Hermitian* (see below) and *positive definite*, which means that

$$\varphi(\overrightarrow{u}, \overrightarrow{u}) > 0, \quad \text{for every } \overrightarrow{u} \neq \overrightarrow{0}.$$

More explicitly, $\varphi\colon E \times E \to \mathbb{R}$ satisfies the following axioms:

$$\begin{aligned}
\varphi(\overrightarrow{u_1} + \overrightarrow{u_2}, \overrightarrow{v}) &= \varphi(\overrightarrow{u_1}, \overrightarrow{v}) + \varphi(\overrightarrow{u_2}, \overrightarrow{v}),\\
\varphi(\overrightarrow{u}, \overrightarrow{v_1} + \overrightarrow{v_2}) &= \varphi(\overrightarrow{u}, \overrightarrow{v_1}) + \varphi(\overrightarrow{u}, \overrightarrow{v_2}),\\
\varphi(\lambda\overrightarrow{u}, \overrightarrow{v}) &= \lambda\varphi(\overrightarrow{u}, \overrightarrow{v}),\\
\varphi(\overrightarrow{u}, \mu\overrightarrow{v}) &= \overline{\mu}\varphi(\overrightarrow{u}, \overrightarrow{v}),\\
\varphi(\overrightarrow{u}, \overrightarrow{v}) &= \overline{\varphi(\overrightarrow{v}, \overrightarrow{u})} \qquad (\varphi \text{ is Hermitian}),\\
\overrightarrow{u} \neq \overrightarrow{0} \text{ implies that } &\varphi(\overrightarrow{u}, \overrightarrow{u}) > 0.
\end{aligned}$$

For simplicity, a map satisfying the above properties is called a *Hermitian positive definite form*. The complex number $\varphi(\overrightarrow{u}, \overrightarrow{v})$ is also called the *inner product of* $\overrightarrow{u}$ *and* $\overrightarrow{v}$. We also define the *quadratic form associated with* φ as the function $\Phi\colon E \to \mathbb{R}_+$ such that

$$\Phi(\overrightarrow{u}) = \varphi(\overrightarrow{u}, \overrightarrow{u}),$$

for all $\overrightarrow{u} \in E$.

Observe that $\varphi(\overrightarrow{0}, \overrightarrow{0}) = 0$, and since φ is positive definite, we have the stronger fact that

$$\varphi(\overrightarrow{u}, \overrightarrow{u}) = 0 \quad \text{iff} \quad \overrightarrow{u} = \overrightarrow{0},$$

that is $\Phi(\overrightarrow{u}) = 0$ iff $\overrightarrow{u} = \overrightarrow{0}$.

The standard example of a Hermitian space is $\mathbb{C}^n$, under the inner product φ defined such that

$$\varphi((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n}.$$

Given a linear map $f\colon E \to E$ (where E is a complex vector space), as in the real case, there is a unique linear map $f^*\colon E \to E$ such that

$$\langle f(\overrightarrow{u}), \overrightarrow{v}\rangle = \langle \overrightarrow{u}, f^*(\overrightarrow{v})\rangle,$$

for all $\overrightarrow{u}, \overrightarrow{v} \in E$. The map f^* is also called the *adjoint of f*.

The inner product $\langle -, - \rangle$ on a Euclidean space E is extended to the Hermitian positive definite form $\langle -, - \rangle_{\mathbb{C}}$ on $E_{\mathbb{C}}$ as follows:

$$\langle \overrightarrow{u_1} + i\overrightarrow{v_1}, \overrightarrow{u_2} + i\overrightarrow{v_2}\rangle_{\mathbb{C}} = \langle \overrightarrow{u_1}, \overrightarrow{u_2}\rangle + \langle \overrightarrow{v_1}, \overrightarrow{v_2}\rangle + i(\langle \overrightarrow{u_2}, \overrightarrow{v_1}\rangle - \langle \overrightarrow{u_1}, \overrightarrow{v_2}\rangle).$$

It is easily verified that $\langle -, - \rangle_{\mathbb{C}}$ is indeed a Hermitian form that is positive definite, and it is clear that $\langle -, - \rangle_{\mathbb{C}}$ agrees with $\langle -, - \rangle$ on real vectors. Then, given any linear map $f\colon E \to E$, it is easily verified that the map $f_{\mathbb{C}}^*$ defined such that

$$f_{\mathbb{C}}^*(\overrightarrow{u} + i\overrightarrow{v}) = f_{\mathbb{C}}^*(\overrightarrow{u}) + i f_{\mathbb{C}}^*(\overrightarrow{v})$$

for all $\overrightarrow{u}, \overrightarrow{v} \in E$ is the adjoint of $f_{\mathbb{C}}$ w.r.t. $\langle -, - \rangle_{\mathbb{C}}$.

Assuming again that E is a Hermitian space, observe that Lemma 2.3 also holds. We have the following crucial lemma relating the eigenvalues of f and f^*.

Lemma 2.6 *Given a Hermitian space E, for any normal linear map $f\colon E \to E$, a vector $\overrightarrow{u}$ is an eigenvector of f for the eigenvalue λ (in $\mathbb{C}$) iff $\overrightarrow{u}$ is an eigenvector of f^* for the eigenvalue $\overline{\lambda}$.*

Proof. First, it is immediately verified that the adjoint of $f - \lambda\,\mathrm{id}$ is $f^* - \overline{\lambda}\,\mathrm{id}$. Furthermore, $f - \lambda\,\mathrm{id}$ is normal. Indeed,

$$\begin{aligned}
(f - \lambda\,\mathrm{id}) \circ (f - \lambda\,\mathrm{id})^* &= (f - \lambda\,\mathrm{id}) \circ (f^* - \overline{\lambda}\,\mathrm{id}),\\
&= f \circ f^* - \overline{\lambda} f - \lambda f^* + \lambda\overline{\lambda}\,\mathrm{id},\\
&= f^* \circ f - \lambda f^* - \overline{\lambda} f + \overline{\lambda}\lambda\,\mathrm{id},\\
&= (f^* - \overline{\lambda}\,\mathrm{id}) \circ (f - \lambda\,\mathrm{id}),\\
&= (f - \lambda\,\mathrm{id})^* \circ (f - \lambda\,\mathrm{id}).
\end{aligned}$$

Applying lemma 2.3 to $f - \lambda\,\mathrm{id}$, for every nonnull vector $\overrightarrow{u}$, we see that

$$(f - \lambda\,\mathrm{id})(\overrightarrow{u}) = \overrightarrow{0} \quad \text{iff} \quad (f^* - \overline{\lambda}\,\mathrm{id})(\overrightarrow{u}) = \overrightarrow{0},$$

which is exactly the statement of the lemma. $\square$

The next lemma shows a very important property of normal linear maps: eigenvectors corresponding to distinct eigenvalues are orthogonal.

Lemma 2.7 *Given a Hermitian space E, for any normal linear map $f\colon E \to E$, if $\overrightarrow{u}$ and $\overrightarrow{v}$ are eigenvectors of f associated with the eigenvalues λ and μ (in $\mathbb{C}$) where $\lambda \neq \mu$, then $\langle \overrightarrow{u}, \overrightarrow{v} \rangle = 0$.*

Proof. Let us compute $\langle f(\overrightarrow{u}), \overrightarrow{v} \rangle$ in two different ways. Since $\overrightarrow{v}$ is an eigenvector of f for μ, by lemma 2.6, $\overrightarrow{v}$ is also an eigenvector of f^* for $\overline{\mu}$, and we have

$$\langle f(\overrightarrow{u}), \overrightarrow{v} \rangle = \langle \lambda \overrightarrow{u}, \overrightarrow{v} \rangle = \lambda \langle \overrightarrow{u}, \overrightarrow{v} \rangle,$$

and

$$\langle f(\overrightarrow{u}), \overrightarrow{v} \rangle = \langle \overrightarrow{u}, f^*(\overrightarrow{v}) \rangle = \langle \overrightarrow{u}, \overline{\mu}\, \overrightarrow{v} \rangle = \mu \langle \overrightarrow{u}, \overrightarrow{v} \rangle,$$

where the last identity holds because of the semi-linearity in the second argument, and thus

$$\lambda \langle \overrightarrow{u}, \overrightarrow{v} \rangle = \mu \langle \overrightarrow{u}, \overrightarrow{v} \rangle,$$

that is

$$(\lambda - \mu) \langle \overrightarrow{u}, \overrightarrow{v} \rangle = 0,$$

which implies that $\langle \overrightarrow{u}, \overrightarrow{v} \rangle = 0$ since $\lambda \neq \mu$. $\square$

We can also show easily that the eigenvalues of a self-adjoint linear map are real.

Lemma 2.8 *Given a Hermitian space E, the eigenvalues of any self-adjoint linear map $f\colon E \to E$ are real.*

Proof. Let z (in $\mathbb{C}$) be an eigenvalue of f and let $\overrightarrow{u}$ be an eigenvector for z. We compute $\langle f(\overrightarrow{u}), \overrightarrow{u} \rangle$ in two different ways. We have

$$\langle f(\overrightarrow{u}), \overrightarrow{u} \rangle = \langle z \overrightarrow{u}, \overrightarrow{u} \rangle = z \langle \overrightarrow{u}, \overrightarrow{u} \rangle,$$

and since $f = f^*$, we also have

$$\langle f(\overrightarrow{u}), \overrightarrow{u} \rangle = \langle \overrightarrow{u}, f^*(\overrightarrow{u}) \rangle = \langle \overrightarrow{u}, f(\overrightarrow{u}) \rangle = \langle \overrightarrow{u}, z \overrightarrow{u} \rangle = \overline{z} \langle \overrightarrow{u}, \overrightarrow{u} \rangle.$$

Thus,

$$z \langle \overrightarrow{u}, \overrightarrow{u} \rangle = \overline{z} \langle \overrightarrow{u}, \overrightarrow{u} \rangle,$$

which implies that $z = \overline{z}$ since $\overrightarrow{u} \neq \overrightarrow{0}$, and z is indeed real. $\square$

Given any subspace W of a Hermitian space E, recall that the *orthogonal* $W^{\perp}$ of W is the subspace defined such that

$$W^{\perp} = \{ \overrightarrow{u} \in E \mid \langle \overrightarrow{u}, \overrightarrow{w} \rangle = 0, \text{ for all } \overrightarrow{w} \in W \}.$$

Recall that $E = W \oplus W^{\perp}$ (this can be easily shown for example by constructing an orthonormal basis of E using the Gram–Schmidt orthonormalization procedure). The same result also holds for Euclidean spaces. The following lemma provides the key to the induction that will allow us to show that a normal linear map can diagonalized. It actually holds for any linear map. We found the inspiration for this Lemma in Berger [3].

Lemma 2.9 *Given a Hermitian space E, for any linear map $f\colon E \to E$, if W is any subspace of E such that $f(W) \subseteq W$ and $f^*(W) \subseteq W$, then $f(W^\perp) \subseteq W^\perp$ and $f^*(W^\perp) \subseteq W^\perp$.*

Proof. If $\overrightarrow{u} \in W^\perp$, then
$$\langle \overrightarrow{u}, \overrightarrow{w} \rangle = 0$$
for all $\overrightarrow{w} \in W$. However,
$$\langle f(\overrightarrow{u}), \overrightarrow{w} \rangle = \langle \overrightarrow{u}, f^*(\overrightarrow{w}) \rangle,$$
and since $f^*(W) \subseteq W$, we have $f^*(\overrightarrow{w}) \in W$, and since $\overrightarrow{u} \in W^\perp$, we get
$$\langle \overrightarrow{u}, f^*(\overrightarrow{w}) \rangle = 0,$$
which shows that
$$\langle f(\overrightarrow{u}), \overrightarrow{w} \rangle = 0$$
for all $\overrightarrow{w} \in W$, that is, $f(\overrightarrow{u}) \in W^\perp$. Thus, $f(W^\perp) \subseteq W^\perp$. The proof that $f^*(W^\perp) \subseteq W^\perp$ is analogous. $\square$

The above Lemma also holds for Euclidean spaces. Although we are ready to prove that for every normal linar map f (over a Hermitian space) there is an orthonormal basis of eigenvectors, we now return to real Euclidean spaces.

If $f\colon E \to E$ is a linear map and $\overrightarrow{w} = \overrightarrow{u} + i\overrightarrow{v}$ is an eigenvector of $f_{\mathbb{C}}\colon E_{\mathbb{C}} \to E_{\mathbb{C}}$ for the eigenvalue $z = \lambda + i\mu$, where $\overrightarrow{u}, \overrightarrow{v} \in E$ and $\lambda, \mu \in \mathbb{R}$, since
$$f_{\mathbb{C}}(\overrightarrow{u} + i\overrightarrow{v}) = f(\overrightarrow{u}) + if(\overrightarrow{v})$$
and
$$f_{\mathbb{C}}(\overrightarrow{u} + i\overrightarrow{v}) = (\lambda + i\mu)(\overrightarrow{u} + i\overrightarrow{v}) = \lambda\overrightarrow{u} - \mu\overrightarrow{v} + i(\mu\overrightarrow{u} + \lambda\overrightarrow{v}),$$
we have
$$f(\overrightarrow{u}) = \lambda\overrightarrow{u} - \mu\overrightarrow{v} \quad \text{and} \quad f(\overrightarrow{v}) = \mu\overrightarrow{u} + \lambda\overrightarrow{v},$$
from which we immediately obtain
$$f_{\mathbb{C}}(\overrightarrow{u} - i\overrightarrow{v}) = (\lambda - i\mu)(\overrightarrow{u} - i\overrightarrow{v}),$$
which shows that $\overline{\overrightarrow{w}} = \overrightarrow{u} - i\overrightarrow{v}$ is an eigenvector of f for $\overline{z} = \lambda - i\mu$. Using this fact, we can prove the following lemma.

Lemma 2.10 *Given a Euclidean space E, for any normal linear map $f\colon E \to E$, if $\overrightarrow{w} = \overrightarrow{u} + i\overrightarrow{v}$ is an eigenvector of $f_{\mathbb{C}}$ associated with the eigenvalue $z = \lambda + i\mu$ (where $\overrightarrow{u}, \overrightarrow{v} \in E$ and $\lambda, \mu \in \mathbb{R}$), if $\mu \neq 0$ (i.e., z is not real) then $\langle \overrightarrow{u}, \overrightarrow{v} \rangle = 0$ and $\langle \overrightarrow{u}, \overrightarrow{u} \rangle = \langle \overrightarrow{v}, \overrightarrow{v} \rangle$, which implies that $\overrightarrow{u}$ and $\overrightarrow{v}$ are linearly independent, and if W is the subspace spanned by $\overrightarrow{u}$ and $\overrightarrow{v}$, then $f(W) = W$ and $f^*(W) = W$. Furthermore, with respect to the (orthogonal) basis $(\overrightarrow{u}, \overrightarrow{v})$, the restriction of f to W has the matrix*

$$\begin{pmatrix} \lambda & \mu \\ -\mu & \lambda \end{pmatrix}.$$

If $\mu = 0$, then λ is a real eigenvalue of f and either $\overrightarrow{u}$ or $\overrightarrow{v}$ is an eigenvector of f for λ. If W is the subspace spanned by $\overrightarrow{u}$ if $\overrightarrow{u} \neq \overrightarrow{0}$, or spanned by $\overrightarrow{v} \neq \overrightarrow{0}$ if $\overrightarrow{u} = \overrightarrow{0}$, then $f(W) \subseteq W$ and $f^*(W) \subseteq W$.

Proof. Since $\overrightarrow{w} = \overrightarrow{u} + i\overrightarrow{v}$ is an eigenvector of $f_{\mathbb{C}}$, by definition, it is nonnull, and either $\overrightarrow{u} \neq \overrightarrow{0}$ or $\overrightarrow{v} \neq \overrightarrow{0}$. From the fact stated just before Lemma 2.10, $\overrightarrow{u} - i\overrightarrow{v}$ is an eigenvector of f for $\lambda - i\mu$. However, if $\mu \neq 0$ then $\lambda + i\mu \neq \lambda - i\mu$, and from Lemma 2.7, the vectors $\overrightarrow{u} + i\overrightarrow{v}$ and $\overrightarrow{u} - i\overrightarrow{v}$ are orthogonal w.r.t. $\langle -, - \rangle_{\mathbb{C}}$, that is,

$$\langle \overrightarrow{u} + i\overrightarrow{v}, \overrightarrow{u} - i\overrightarrow{v} \rangle_{\mathbb{C}} = \langle \overrightarrow{u}, \overrightarrow{u} \rangle - \langle \overrightarrow{v}, \overrightarrow{v} \rangle + 2i\langle \overrightarrow{u}, \overrightarrow{v} \rangle = 0.$$

Thus, we get $\langle \overrightarrow{u}, \overrightarrow{v} \rangle = 0$ and $\langle \overrightarrow{u}, \overrightarrow{u} \rangle = \langle \overrightarrow{v}, \overrightarrow{v} \rangle$, and since $\overrightarrow{u} \neq \overrightarrow{0}$ or $\overrightarrow{v} \neq \overrightarrow{0}$, $\overrightarrow{u}$ and $\overrightarrow{v}$ are linearly independent. Since

$$f(\overrightarrow{u}) = \lambda\overrightarrow{u} - \mu\overrightarrow{v} \quad \text{and} \quad f(\overrightarrow{v}) = \mu\overrightarrow{u} + \lambda\overrightarrow{v}$$

and since by lemma 2.6, $\overrightarrow{u} + i\overrightarrow{v}$ is an eigenvector of f^* for $\lambda - i\mu$, we have

$$f^*(\overrightarrow{u}) = \lambda\overrightarrow{u} + \mu\overrightarrow{v} \quad \text{and} \quad f^*(\overrightarrow{v}) = -\mu\overrightarrow{u} + \lambda\overrightarrow{v},$$

and thus $f(W) = W$ and $f^*(W) = W$, where W is the subspace spanned by $\overrightarrow{u}$ and $\overrightarrow{v}$.

When $\mu = 0$, we have

$$f(\overrightarrow{u}) = \lambda\overrightarrow{u} \quad \text{and} \quad f(\overrightarrow{v}) = \lambda\overrightarrow{v},$$

and since $\overrightarrow{u} \neq \overrightarrow{0}$ or $\overrightarrow{v} \neq \overrightarrow{0}$, either $\overrightarrow{u}$ or $\overrightarrow{v}$ is an eigenvector of f for λ. If W is the subspace spanned by $\overrightarrow{u}$ if $\overrightarrow{u} \neq \overrightarrow{0}$, or spanned by $\overrightarrow{v}$ if $\overrightarrow{u} = \overrightarrow{0}$, it is obvious that $f(W) \subseteq W$ and $f^*(W) \subseteq W$. Note that $\lambda = 0$ is possible, and this is why $\subseteq$ can't be replaced by $=$. $\square$

The beginning of the proof of Lemma 2.10 actually shows that for every linear map $f\colon E \to E$, there is some subspace W such $f(W) \subseteq W$, where W has dimension 1 or 2. In general, it doesn't seem possible to prove that $W^{\perp}$ is invariant under f. However, this happens when f is normal, and in this case, other nice things also happen. We can finally prove our first main theorem.

Theorem 2.11 *Given a Euclidean space E of dimension n, for every normal linear map $f\colon E \to E$, there is an orthonormal basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ such that the matrix of f w.r.t. this basis is a block diagonal matrix of the form*

$$\begin{pmatrix} A_1 & & \dots & \\ & A_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & A_p \end{pmatrix}$$

such that each block A_i is either a one-dimensional matrix (i.e., a real scalar) or a two-dimensional matrix of the form

$$A_i = \begin{pmatrix} \lambda_i & -\mu_i \\ \mu_i & \lambda_i \end{pmatrix}$$

where $\lambda_i, \mu_i \in \mathbb{R}$, with $\mu_i > 0$.

Proof. We proceed by induction on the dimension n of E as follows. If $n = 1$, the result is trivial. Assume now that $n \geq 2$. First, since $\mathbb{C}$ is algebraically closed (i.e., every polynomial has a root in $\mathbb{C}$), the linear map $f_{\mathbb{C}}\colon E_{\mathbb{C}} \to E_{\mathbb{C}}$ has some eigenvalue $z = \lambda + i\mu$ (where $\lambda, \mu \in \mathbb{R}$). Let $\overrightarrow{w} = \overrightarrow{u} + i\overrightarrow{v}$ be some eigenvector of $f_{\mathbb{C}}$ for $\lambda + i\mu$ (where $\overrightarrow{u}, \overrightarrow{v} \in E$). We can now apply Lemma 2.10.

If $\mu = 0$, then either $\overrightarrow{u}$ or $\overrightarrow{v}$ is an eigenvector of f for $\lambda \in \mathbb{R}$. Let W be the subspace of dimension 1 spanned by $\overrightarrow{e_1} = \frac{\overrightarrow{u}}{\|\overrightarrow{u}\|}$ if $\overrightarrow{u} \neq \overrightarrow{0}$, or by $\overrightarrow{e_1} = \frac{\overrightarrow{v}}{\|\overrightarrow{v}\|}$ otherwise. It is obvious that $f(W) \subseteq W$ and $f^*(W) \subseteq W$. The orthogonal $W^{\perp}$ of W has dimension $n - 1$, and by Lemma 2.9, we have $f(W^{\perp}) \subseteq W^{\perp}$. But the restriction of f to $W^{\perp}$ is also normal, and we conclude by applying the induction hypothesis to $W^{\perp}$.

If $\mu \neq 0$, then $\langle u, v \rangle = 0$ and $\langle u, u \rangle = \langle v, v \rangle$, and if W is the subspace spanned by $u/\|u\|$ and $v/\|v\|$, then $f(W) = W$ and $f^*(W) = W$. We also know that the restriction of f to W has the matrix

$$\begin{pmatrix} \lambda & \mu \\ -\mu & \lambda \end{pmatrix}.$$

with respect to the basis $(u/\|u\|, v/\|v\|)$. If $\mu < 0$, we let $\lambda_1 = \lambda$, $\mu_1 = -\mu$, $e_1 = u/\|u\|$, and $e_2 = v/\|v\|$. If $\mu > 0$, we let $\lambda_1 = \lambda$, $\mu_1 = \mu$, $e_1 = v/\|v\|$, and $e_2 = u/\|u\|$. In all cases, it is easily verified that the matrix of the restriction of f to W w.r.t. the orthonormal basis (e_1, e_2) is

$$A_1 = \begin{pmatrix} \lambda_1 & -\mu_1 \\ \mu_1 & \lambda_1 \end{pmatrix},$$

where $\lambda_1, \mu_1 \in \mathbb{R}$, with $\mu_1 > 0$. However, $W^{\perp}$ has dimension $n - 2$, and by Lemma 2.9, $f(W^{\perp}) \subseteq W^{\perp}$. Since the restriction of f to $W^{\perp}$ is also normal, we conclude by applying the induction hypothesis to $W^{\perp}$. $\square$

After this relatively hard work, we can easily obtain some nice normal forms for the matrices of self-adjoint, skew self-adjoint, and orthogonal, linear maps. However, for the sake of completeness (and since we have all the tools to so do), we go back to the case of a Hermitian space and show that normal linear maps can be diagonalized with respect to an orthonormal basis.

Theorem 2.12 *Given a Hermitian space E of dimension n, for every normal linear map $f: E \to E$, there is an orthonormal basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ of eigenvectors of f such that the matrix of f w.r.t. this basis is a diagonal matrix*

$$\begin{pmatrix} \lambda_1 & & \ldots & \\ & \lambda_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \lambda_n \end{pmatrix}$$

where $\lambda_i \in \mathbb{C}$.

Proof. We proceed by induction on the dimension n of E as follows. If $n = 1$, the result is trivial. Assume now that $n \geq 2$. Since $\mathbb{C}$ is algebraically closed (i.e., every polynomial has a root in $\mathbb{C}$), the linear map $f: E \to E$ has some eigenvalue $\lambda \in \mathbb{C}$, and let $\overrightarrow{w}$ be some eigenvector for λ. Let W be the subspace of dimension 1 spanned by $\overrightarrow{w}$. Clearly, $f(W) \subseteq W$. By Lemma 2.6, $\overrightarrow{w}$ is an eigenvector of f^* for $\overline{\lambda}$, and thus $f^*(W) \subseteq W$. By Lemma 2.9, we also have $f(W^\perp) \subseteq W^\perp$. The restriction of f to $W^\perp$ is still normal, and we conclude by applying the induction hypothesis to $W^\perp$ (whose dimension is $n - 1$). $\square$

Thus, in particular, self-adjoint, skew self-adjoint, and orthogonal, linear maps can be diagonalized with respect to an orthonormal basis of eigenvectors. In this latter case though, an orthogonal map is called a *unitary* map. Also, Lemma 2.8 shows that the eigenvalues of a self-adjoint linear map are real. It is easily shown that skew self-adjoint maps have eigenvalues that are pure imaginary of null, and that unitary maps have eigenvalues of absolute value 1.

Remark: There is a converse to Theorem 2.12, namely, if there is an orthonormal basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ of eigenvectors of f, then f is normal. We leave the easy proof as an exercise.

3 Self-Adjoint, Skew Self-Adjoints, And Orthogonal, Linear Maps

We begin with self-adjoint maps.

Theorem 3.1 *Given a Euclidean space E of dimension n, for every self-adjoint linear map $f: E \to E$, there is an orthonormal basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ of eigenvectors of f such that the matrix of f w.r.t. this basis is a diagonal matrix*

$$\begin{pmatrix} \lambda_1 & & \dots & \\ & \lambda_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & \lambda_n \end{pmatrix}$$

where $\lambda_i \in \mathbb{R}$.

Proof. The case $n = 1$ is trivial. If $n \geq 2$, we need to show that $f: E \to E$ has some real eigenvalue. There are several ways to do so. One method is to observe that the linear map $f_{\mathbb{C}}: E_{\mathbb{C}} \to E_{\mathbb{C}}$ is also self-adjoint, and by Lemma 2.8, the eigenvalues of $f_{\mathbb{C}}$ are all real. This implies that f itself has some real eigenvalue, and in fact, all eigenvalues of f are real. We now give a more direct method not involving the complexification of $\langle -, - \rangle$ and Lemma 2.8.

Since $\mathbb{C}$ is algebraically closed, $f_{\mathbb{C}}$ has some eigenvalue $\lambda + i\mu$, and let $\overrightarrow{u} + i\overrightarrow{v}$ be some eigenvector of $f_{\mathbb{C}}$ for $\lambda + i\mu$, where $\lambda, \mu \in \mathbb{R}$ and $\overrightarrow{u}, \overrightarrow{v} \in E$. We saw in the proof of Lemma 2.10 that

$$f(\overrightarrow{u}) = \lambda\overrightarrow{u} - \mu\overrightarrow{v} \quad \text{and} \quad f(\overrightarrow{v}) = \mu\overrightarrow{u} + \lambda\overrightarrow{v}.$$

Since $f = f^*$,

$$\langle f(\overrightarrow{u}), \overrightarrow{v} \rangle = \langle \overrightarrow{u}, f(\overrightarrow{v}) \rangle$$

for all $\overrightarrow{u}, \overrightarrow{v} \in E$. Applying this to

$$f(\overrightarrow{u}) = \lambda\overrightarrow{u} - \mu\overrightarrow{v} \quad \text{and} \quad f(\overrightarrow{v}) = \mu\overrightarrow{u} + \lambda\overrightarrow{v},$$

we get

$$\langle f(\overrightarrow{u}), \overrightarrow{u} \rangle = \langle \lambda\overrightarrow{u} - \mu\overrightarrow{v}, \overrightarrow{v} \rangle = \lambda\langle \overrightarrow{u}, \overrightarrow{v} \rangle - \mu\langle \overrightarrow{v}, \overrightarrow{v} \rangle$$

and

$$\langle \overrightarrow{u}, f(\overrightarrow{v}) \rangle = \langle \overrightarrow{u}, \mu\overrightarrow{u} + \lambda\overrightarrow{v} \rangle = \mu\langle \overrightarrow{u}, \overrightarrow{u} \rangle + \lambda\langle \overrightarrow{u}, \overrightarrow{v} \rangle,$$

and thus we get

$$\lambda\langle \overrightarrow{u}, \overrightarrow{v} \rangle - \mu\langle \overrightarrow{v}, \overrightarrow{v} \rangle = \mu\langle \overrightarrow{u}, \overrightarrow{u} \rangle + \lambda\langle \overrightarrow{u}, \overrightarrow{v} \rangle,$$

that is

$$\mu(\langle \overrightarrow{u}, \overrightarrow{u} \rangle + \langle \overrightarrow{v}, \overrightarrow{v} \rangle) = 0,$$

which implies $\mu = 0$ since either $\overrightarrow{u} \neq \overrightarrow{0}$ or $\overrightarrow{v} \neq \overrightarrow{0}$. Therefore, λ is a real eigenvalue of f.

Now, going back to the proof of Theorem 2.11, only the case where $\mu = 0$ applies, and the induction shows that all the blocks are one-dimensional. $\square$

Theorem 3.1 implies that if $\lambda_1, \dots, \lambda_p$ are the distinct real eigenvalues of f and E_i is the eigenspace associated with λ_i, then

$$E = E_1 \oplus \cdots \oplus E_p,$$

where E_i and E_j are orthogonal for all $i \neq j$.

Remark: Another way to prove that a self-adjoint map has a real eigenvalue is to use a little bit of calculus. We learned such a proof from Herman Gluck. The idea is to consider the real-valued function $\Phi: E \to \mathbb{R}$ defined such that

$$\Phi(\overrightarrow{u}) = \langle f(\overrightarrow{u}), \overrightarrow{u} \rangle$$

for every $\overrightarrow{u} \in E$. This function is C^∞, and if we represent f by a matrix A over some orthonormal basis, it is easy to compute the gradient vector

$$\nabla\Phi(X) = \left(\frac{\partial\Phi}{\partial x_1}(X), \ldots, \frac{\partial\Phi}{\partial x_n}(X)\right)$$

of Φ at X. Indeed, we find that

$$\nabla\Phi(X) = (A + A^\top)(X),$$

where X is a column vector of size n. But since f is self-adjoint, $A = A^\top$, and thus

$$\nabla\Phi(X) = 2A(X).$$

The next step is to find the maximum of the function Φ on the sphere

$$S^{n-1} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1^2 + \cdots + x_n^2 = 1\}.$$

Since S^{n-1} is compact and Φ is continuous, and in fact C^∞, Φ takes a maximum at some X on S^{n-1}. But then, it is well know that at an extremum X of Φ, we must have

$$d\Phi_X(Y) = \langle \nabla\Phi(X), Y \rangle = 0$$

for all tangent vectors Y to S^{n-1} at X, and so, $\nabla\Phi(X)$ is orthogonal to the tangent plane at X, which means that

$$\nabla\Phi(X) = \lambda X$$

for some $\lambda \in \mathbb{R}$. Since $\nabla\Phi(X) = 2A(X)$, we get

$$2A(X) = \lambda X,$$

and thus $\lambda/2$ is a real eigenvalue of A (i.e. of f).

Next, we consider skew self-adjoint maps.

Theorem 3.2 *Given a Euclidean space E of dimension n, for every skew self-adjoint linear map $f: E \to E$, there is an orthonormal basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ such that the matrix of f w.r.t. this basis is a block diagonal matrix of the form*

$$\begin{pmatrix} A_1 & & \dots & \\ & A_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & A_p \end{pmatrix}$$

such that each block A_i is either 0 *or a two-dimensional matrix of the form*

$$A_i = \begin{pmatrix} 0 & -\mu_i \\ \mu_i & 0 \end{pmatrix}$$

where $\mu_i \in \mathbb{R}$, with $\mu_i > 0$. In particular, the eigenvalues of $f_{\mathbb{C}}$ are pure imaginary of the form $i\mu_i$, or 0.

Proof. The case where $n = 1$ is trivial. As in the proof of Theorem 2.11, $f_{\mathbb{C}}$ has some eigenvalue $z = \lambda + i\mu$, where $\lambda, \mu \in \mathbb{R}$. We claim that $\lambda = 0$. First, we show that

$$\langle f(\overrightarrow{w}), \overrightarrow{w} \rangle = 0$$

for all $\overrightarrow{w} \in E$. Indeed, since $f = -f^*$, we get

$$\langle f(\overrightarrow{w}), \overrightarrow{w} \rangle = \langle \overrightarrow{w}, f^*(\overrightarrow{w}) \rangle = \langle \overrightarrow{w}, -f(\overrightarrow{w}) \rangle = -\langle \overrightarrow{w}, f(\overrightarrow{w}) \rangle = -\langle f(\overrightarrow{w}), \overrightarrow{w} \rangle,$$

since $\langle -, - \rangle$ is symmetric. This implies that

$$\langle f(\overrightarrow{w}), \overrightarrow{w} \rangle = 0.$$

Applying this to $\overrightarrow{u}$ and $\overrightarrow{v}$ and using the fact that

$$f(\overrightarrow{u}) = \lambda \overrightarrow{u} - \mu \overrightarrow{v} \quad \text{and} \quad f(\overrightarrow{v}) = \mu \overrightarrow{u} + \lambda \overrightarrow{v},$$

we get

$$0 = \langle f(\overrightarrow{u}), \overrightarrow{u} \rangle = \langle \lambda \overrightarrow{u} - \mu \overrightarrow{v}, \overrightarrow{u} \rangle = \lambda \langle \overrightarrow{u}, \overrightarrow{u} \rangle - \mu \langle \overrightarrow{u}, \overrightarrow{v} \rangle$$

and

$$0 = \langle f(\overrightarrow{v}), \overrightarrow{v} \rangle = \langle \mu \overrightarrow{u} + \lambda \overrightarrow{v}, \overrightarrow{v} \rangle = \mu \langle \overrightarrow{u}, \overrightarrow{v} \rangle + \lambda \langle \overrightarrow{v}, \overrightarrow{v} \rangle,$$

from which, by addition, we get

$$\lambda(\langle \overrightarrow{v}, \overrightarrow{v} \rangle + \langle \overrightarrow{v}, \overrightarrow{v} \rangle) = 0.$$

Since $\overrightarrow{u} \neq \overrightarrow{0}$ or $\overrightarrow{v} \neq \overrightarrow{0}$, we have $\lambda = 0$.

Then, going back to the proof of Theorem 2.11, unless $\mu = 0$, the case where $\overrightarrow{u}$ and $\overrightarrow{v}$ are orthogonal and span a subspace of dimension 2 applies, and the induction shows that all the blocks are two-dimensional or reduced to 0. □

Remark: One will note that if f is skew self-adjoint, then $if_{\mathbb{C}}$ is self-adjoint w.r.t. $\langle -, - \rangle_{\mathbb{C}}$. By Lemma 2.8, the map $if_{\mathbb{C}}$ has real eigenvalues, which implies that the eigenvalues of $f_{\mathbb{C}}$ are pure imaginary or 0.

Finally, we consider orthogonal linear maps.

Theorem 3.3 *Given a Euclidean space E of dimension n, for every orthogonal linear map $f: E \to E$, there is an orthonormal basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ such that the matrix of f w.r.t. this basis is a block diagonal matrix of the form*

$$\begin{pmatrix} A_1 & & \ldots & \\ & A_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & A_p \end{pmatrix}$$

such that each block A_i is either 1, -1, or a two-dimensional matrix of the form

$$A_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

In particular, the eigenvalues of $f_{\mathbb{C}}$ are of the form $\cos\theta_i \pm i\sin\theta_i$, with $0 < \theta_i < \pi$, or 1, or -1.

Proof. The case where $n = 1$ is trivial. As in the proof of Theorem 2.11, $f_{\mathbb{C}}$ has some eigenvalue $z = \lambda + i\mu$, where $\lambda, \mu \in \mathbb{R}$. Since $f \circ f^* = f^* \circ f = \mathrm{id}$, the map f is invertible. In fact, the eigenvalues of f have absolute value 1. Indeed, if z (in $\mathbb{C}$) is an eigenvalue of f and $\overrightarrow{u}$ is an eigenvector for z, we have

$$\langle f(\overrightarrow{u}), f(\overrightarrow{u}) \rangle = \langle z\overrightarrow{u}, z\overrightarrow{u} \rangle = z\overline{z}\langle \overrightarrow{u}, \overrightarrow{u} \rangle$$

and

$$\langle f(\overrightarrow{u}), f(\overrightarrow{u}) \rangle = \langle \overrightarrow{u}, (f^* \circ f)(\overrightarrow{u}) \rangle = \langle \overrightarrow{u}, \overrightarrow{u} \rangle,$$

from which we get

$$z\overline{z}\langle \overrightarrow{u}, \overrightarrow{u} \rangle = \langle \overrightarrow{u}, \overrightarrow{u} \rangle.$$

Since $\overrightarrow{u} \neq \overrightarrow{0}$, we have $z\overline{z} = 1$, i.e. $|z| = 1$. As a consequence, the eigenvalues of $f_{\mathbb{C}}$ are either of the form $\cos\theta \pm i\sin\theta$, or 1, or -1. The theorem then follows immediately from Theorem 2.11, where the condition $\mu > 0$ implies that $\sin\theta_i > 0$, and thus, $0 < \theta_i < \pi$. $\square$

If f is orthogonal and $\det(f) = +1$ (f is a rotation), the number of -1 must be even, and these entries can be paired to form two-dimensional blocks.

The theorems of this section and of the previous section can be immediately applied to matrices.

4 Normal, Symmetric, Skew Symmetric, Orthogonal, Hermitian, Skew Hermitian, and Unitary, Matrices

First, we consider real matrices. Recall the following definitions.

Definition 4.1 Given a real $m \times n$ matrix A, the *transpose* $A^\top$ *of* A is the $n \times m$ matrix $A^\top = (a^\top_{i,j})$ defined such that

$$a^\top_{i,j} = a_{j,i}$$

for all i, j, $1 \leq i \leq m$, $1 \leq j \leq n$. A real $n \times n$ matrix A is

- *normal* if
$$A\,A^\top = A^\top A,$$
- *symmetric* if
$$A^\top = A,$$
- *skew symmetric* if
$$A^\top = -A,$$
- *orthogonal* if
$$A\,A^\top = A^\top A = I_n.$$

It is easily verified that when E is a Euclidean space and $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ is an orthonormal basis for E, if a linear map $f\colon E \to E$ has the matrix A w.r.t. the basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$, then its adjoint f^* has the matrix $A^\top$. Consequently, a normal linear map has a normal matrix, a self-adjoint linear map has a symmetric matrix, a skew self-adjoint linear has a skew symmetric matrix, and an orthogonal linear map has an orthogonal matrix. If E and F are Euclidean spaces, $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ is an orthonormal basis for E and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$ is an orthonormal basis for F, if a linear map $f\colon E \to F$ has the matrix A w.r.t. the bases $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$, then its adjoint f^* has the matrix $A^\top$ w.r.t. the bases $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$ and $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$.

Furthermore, if $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ is another orthonormal basis for E and P is the change of basis matrix whose columns are the components of the $\overrightarrow{u_i}$ w.r.t. the basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$, then P is orthogonal, and for any linear map $f\colon E \to E$, if A is the matrix of f w.r.t $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ and B is the matrix of f w.r.t. $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$, then

$$B = P^\top AP.$$

As a consequence, Theorems 2.11 and 3.1–3.3 can be restated as follows.

Theorem 4.2 *For every normal matrix A, there is an orthogonal matrix P and a block diagonal matrix D such that $A = PD\,P^{\top}$, where D is of the form*

$$D = \begin{pmatrix} D_1 & & \dots & \\ & D_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & D_p \end{pmatrix}$$

such that each block D_i is either a one-dimensional matrix (i.e., a real scalar) or a two-dimensional matrix of the form

$$D_i = \begin{pmatrix} \lambda_i & -\mu_i \\ \mu_i & \lambda_i \end{pmatrix}$$

where $\lambda_i, \mu_i \in \mathbb{R}$, with $\mu_i > 0$.

Theorem 4.3 *For every symmetric matrix A, there is an orthogonal matrix P and a diagonal matrix D such that $A = PD\,P^{\top}$, where D is of the form*

$$D = \begin{pmatrix} \lambda_1 & & \dots & \\ & \lambda_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & \lambda_n \end{pmatrix}$$

where $\lambda_i \in \mathbb{R}$.

Theorem 4.4 *For every skew symmetric matrix A, there is an orthogonal matrix P and a block diagonal matrix D such that $A = PD\,P^{\top}$, where D is of the form*

$$D = \begin{pmatrix} D_1 & & \dots & \\ & D_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & D_p \end{pmatrix}$$

such that each block D_i is either 0 or a two-dimensional matrix of the form

$$D_i = \begin{pmatrix} 0 & -\mu_i \\ \mu_i & 0 \end{pmatrix}$$

where $\mu_i \in \mathbb{R}$, with $\mu_i > 0$. In particular, the eigenvalues of A are pure imaginary of the form $i\mu_i$, or 0.

Theorem 4.5 *For every orthogonal matrix A, there is an orthogonal matrix P and a block diagonal matrix D such that $A = PD\,P^\top$, where D is of the form*

$$D = \begin{pmatrix} D_1 & & \dots & \\ & D_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & D_p \end{pmatrix}$$

such that each block D_i is either 1, -1, or a two-dimensional matrix of the form

$$D_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

In particular, the eigenvalues of A are of the form $\cos\theta_i \pm i\sin\theta_i$, with $0 < \theta_i < \pi$, or 1, or -1.

We now consider complex matrices.

Definition 4.6 Given a complex $m\times n$ matrix A, the *transpose $A^\top$ of A* is the $n\times m$ matrix $A^\top = (a^\top_{i,j})$ defined such that

$$a^\top_{i,j} = a_{j,i}$$

for all i, j, $1 \leq i \leq m$, $1 \leq j \leq n$. The *conjugate $\overline{A}$ of A* is the $m \times n$ matrix $\overline{A} = (b_{i,j})$ defined such that

$$b_{i,j} = \overline{a}_{i,j}$$

for all i, j, $1 \leq i \leq m$, $1 \leq j \leq n$. Given an $m \times n$ complex matrix A, the *adjoint A^* of A* is the matrix defined such that

$$A^* = \overline{(A^\top)} = (\overline{A})^\top.$$

A complex $n \times n$ matrix A is

- *normal* if
$$AA^* = A^*A,$$
- *Hermitian* if
$$A^* = A,$$
- *skew Hermitian* if
$$A^* = -A,$$
- *unitary* if
$$AA^* = A^*A = I_n.$$

It is easily verified that when E is a Hermitian space and $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ is an orthonormal basis for E, if a linear map $f\colon E \to E$ has the matrix A w.r.t. the basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$, then its adjoint f^* has the matrix A^*. Consequently, a normal linear map has a normal matrix, a self-adjoint linear map has a Hermitian matrix, a skew self-adjoint linear has a skew Hermitian matrix, and a unitary linear map has a unitary matrix. If E and F are Hermitian spaces, $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ is an orthonormal basis for E and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$ is an orthonormal basis for F, if a linear map $f\colon E \to F$ has the matrix A w.r.t. the bases $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$, then its adjoint f^* has the matrix A^* w.r.t. the bases $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$ and $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$.

Furthermore, if $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ is another orthonormal basis for E and P is the change of basis matrix whose columns are the components of the $\overrightarrow{u_i}$ w.r.t. the basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$, then P is unitary, and for any linear map $f\colon E \to E$, if A is the matrix of f w.r.t $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ and B is the matrix of f w.r.t. $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$, then

$$B = P^*AP.$$

Theorem 2.12 can be restated in terms of matrices as follows. We can also say a little more about eigenvalues (easy exercise left to the reader).

Theorem 4.7 *For every complex normal matrix A, there is a unitary matrix U and a diagonal matrix D such that $A = UDU^*$. Furthermore, if A is Hermitian, D is a real matrix, if A is skew Hermitian, then the entries in D are pure imaginary or null, and if A is unitary, then the entries in D have absolute value* 1.

5 Singular Value Decomposition (SVD), Polar Form

In this section, we assume that we are dealing with a real Euclidean space E. Let $f\colon E \to E$ be any linear map. In general, it may not be possible to diagonalize f. However, note that $f^* \circ f$ is self-adjoint, since

$$\langle (f^* \circ f)(\overrightarrow{u}), \overrightarrow{v} \rangle = \langle f(\overrightarrow{u}), f(\overrightarrow{v}) \rangle = \langle \overrightarrow{u}, (f^* \circ f)(\overrightarrow{v}) \rangle.$$

Similarly, $f \circ f^*$ is self-adjoint.

The fact that $f^* \circ f$ and $f \circ f^*$ are self-adjoint is very important, because it implies that $f^* \circ f$ and $f \circ f^*$ can be diagonalized and that they have real eigenvalues. In fact, these eigenvalues are all ≥ 0. Indeed, if $\overrightarrow{u}$ is an eigenvector of $f^* \circ f$ for the eigenvalue λ, then

$$\langle (f^* \circ f)(\overrightarrow{u}), \overrightarrow{u} \rangle = \langle f(\overrightarrow{u}), f(\overrightarrow{u}) \rangle$$

and

$$\langle (f^* \circ f)(\overrightarrow{u}), \overrightarrow{u} \rangle = \lambda \langle \overrightarrow{u}, \overrightarrow{u} \rangle,$$

and thus

$$\lambda\langle \overrightarrow{u}, \overrightarrow{u}\rangle = \langle f(\overrightarrow{u}), f(\overrightarrow{u})\rangle,$$

which implies that $\lambda \geq 0$, since $\langle -, -\rangle$ is definite positive. A similar proof applies to $f \circ f^*$. Thus, the eigenvalues of $f^* \circ f$ are of the form $\mu_1^2, \ldots, \mu_r^2$ or 0, where $\mu_i > 0$, and similarly for $f \circ f^*$. The situation is even better, since we will show shortly that $f^* \circ f$ and $f \circ f^*$ have the same eigenvalues.

Remark: Given any two linear maps $f\colon E \to F$ and $g\colon F \to E$, where $\dim(E) = n$ and $\dim(F) = m$, it can be shown that

$$(-\lambda)^m \det(g \circ f - \lambda\, I_n) = (-\lambda)^n \det(f \circ g - \lambda\, I_m),$$

and thus $g \circ f$ and $f \circ g$ always have the same nonnull eigenvalues!

The square roots $\mu_i > 0$ of the positive eigenvalues of $f^* \circ f$ (and $f \circ f^*$) are called the *singular values of f*. A self-adjoint linear map $f\colon E \to E$ whose eigenvalues are ≥ 0 is called *positive*, and if f is also invertible, *positive definite*. In the latter case, every eigenvalue is strictly positive. We just showed that $f^* \circ f$ and $f \circ f^*$ are positive self-adjoint linear maps.

The wonderful thing about the singular value decomposition is that there exist two orthonormal bases $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$ such that with respect to these bases, f is a diagonal matrix consisting of the singular values of f, or 0. First, we show some useful relationships between the kernels and the images of f, f^*, $f^* \circ f$, and $f \circ f^*$. Recall that if $f\colon E \to F$ is a linear map, the *image* $\mathrm{Im}\, f$ *of* f is the subspace $f(E)$ of F, and the *rank of* f is the dimension $\dim(\mathrm{Im}\, f)$ of its image. Also recall that

$$\dim(\mathrm{Ker}\, f) + \dim(\mathrm{Im}\, f) = \dim(E),$$

and that for every subspace W of E

$$\dim(W) + \dim(W^{\perp}) = \dim(E).$$

Lemma 5.1 *Given any two Euclidean spaces E and F, where E has dimension n and F has dimension m, for any linear map $f\colon E \to F$, we have*

$$\begin{aligned}
\mathrm{Ker}\, f &= \mathrm{Ker}\,(f^* \circ f),\\
\mathrm{Ker}\, f^* &= \mathrm{Ker}\,(f \circ f^*),\\
\mathrm{Ker}\, f &= (\mathrm{Im}\, f^*)^{\perp},\\
\mathrm{Ker}\, f^* &= (\mathrm{Im}\, f)^{\perp},\\
\dim(\mathrm{Im}\, f) &= \dim(\mathrm{Im}\, f^*),\\
\dim(\mathrm{Ker}\, f) &= \dim(\mathrm{Ker}\, f^*),
\end{aligned}$$

and f, f^*, $f^* \circ f$, and $f \circ f^*$, have the same rank.

Proof. To simplify the notation, we will denote the inner products on E and F by the same symbol $\langle -, - \rangle$ (to avoid subscripts). If $f(\overrightarrow{u}) = \overrightarrow{0}$, then $(f^* \circ f)(\overrightarrow{u}) = f^*(f(\overrightarrow{u})) = f^*(\overrightarrow{0}) = \overrightarrow{0}$, and so $\operatorname{Ker} f \subseteq \operatorname{Ker}(f^* \circ f)$. By definition of f^*, we have

$$\langle f(\overrightarrow{u}), f(\overrightarrow{u}) \rangle = \langle (f^* \circ f)(\overrightarrow{u}), \overrightarrow{u} \rangle$$

for all $\overrightarrow{u} \in E$. If $(f^* \circ f)(\overrightarrow{u}) = \overrightarrow{0}$, since $\langle -, - \rangle$ is positive definite, we must have $f(\overrightarrow{u}) = \overrightarrow{0}$, and so $\operatorname{Ker}(f^* \circ f) \subseteq \operatorname{Ker} f$. Therefore,

$$\operatorname{Ker} f = \operatorname{Ker}(f^* \circ f).$$

The proof that $\operatorname{Ker} f^* = \operatorname{Ker}(f \circ f^*)$ is similar.

By definition of f^*, we have

$$\langle f(\overrightarrow{u}), \overrightarrow{v} \rangle = \langle \overrightarrow{u}, f^*(\overrightarrow{v}) \rangle$$

for all $\overrightarrow{u} \in E$ and all $\overrightarrow{v} \in F$. This immediately implies that

$$\operatorname{Ker} f = (\operatorname{Im} f^*)^{\perp} \quad \text{and} \quad \operatorname{Ker} f^* = (\operatorname{Im} f)^{\perp}.$$

Since

$$\dim(\operatorname{Im} f) = n - \dim(\operatorname{Ker} f)$$

and

$$\dim((\operatorname{Im} f^*)^{\perp}) = n - \dim(\operatorname{Im} f^*),$$

from

$$\operatorname{Ker} f = (\operatorname{Im} f^*)^{\perp}$$

we also have

$$\dim(\operatorname{Ker} f) = \dim((\operatorname{Im} f^*)^{\perp}),$$

from which we obtain

$$\dim(\operatorname{Im} f) = \dim(\operatorname{Im} f^*).$$

The above immediately implies that $\dim(\operatorname{Ker} f) = \dim(\operatorname{Ker} f^*)$. From all this, we easily deduce that

$$\dim(\operatorname{Im} f) = \dim(\operatorname{Im}(f^* \circ f)) = \dim(\operatorname{Im}(f \circ f^*)),$$

i.e., f, f^*, $f^* \circ f$, and $f \circ f^*$, have the same rank. $\square$

The next Lemma shows a very useful property of positive self-adjoint linear maps.

Lemma 5.2 *Given a Euclidean space E of dimension n, for any positive self-adjoint linear map $f\colon E \to E$, there is a unique positive self-adjoint linear map $h\colon E \to E$ such that $f = h^2 = h \circ h$. Furthermore, $\operatorname{Ker} f = \operatorname{Ker} h$, and if $\mu_1, \ldots, \mu_p$ are the distinct eigenvalues of h and E_i is the eigenspace associated with μ_i, then $\mu_1^2, \ldots, \mu_p^2$ are the distinct eigenvalues of f, and E_i is the the eigenspace associated with μ_i^2,* *Proof*. Since f is self-adjoint, by Theorem 3.1, there is an orthonormal basis $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ consisting of eigenvectors of f, and if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of f, we know that $\lambda_i \in \mathbb{R}$. Since f is assumed to be positive, we have $\lambda_i \geq 0$, and we can write $\lambda_i = \mu_i^2$, where $\mu_i \geq 0$. If we define $h: E \rightarrow E$ by its action on the basis $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$, so that

$$h(\overrightarrow{u_i}) = \mu_i \overrightarrow{u_i},$$

it is obvious that $f = h^2$ and that h is positive self-adjoint (since its matrix over the orthonormal basis $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ is diagonal, thus symmetric). It remains to prove that h is uniquely determined by f. Let $g: E \rightarrow E$ be any positive self-adjoint linear map such that $f = g^2$. Then, there is an orthonormal basis $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$ of eigenvectors of g, and let $\mu_1, \ldots, \mu_n$ be the eigenvalues of g, where $\mu_i \geq 0$. Note that

$$f(\overrightarrow{v_i}) = g^2(\overrightarrow{v_i}) = g(g(\overrightarrow{v_i})) = \mu_i^2 \overrightarrow{v_i},$$

so that $\overrightarrow{v_i}$ is an eigenvector of f for the eigenvalue μ_i^2. If $\mu_1, \ldots \mu_p$ are the distinct eigenvalues of g and $E_1, \ldots, E_p$ are the corresponding eigenspaces, the above argument shows that each E_i is a subspace of the eigenspace U_i of f associated with μ_i^2. However, we observed (just after Theorem 3.1) that

$$E = E_1 \oplus \cdots \oplus E_p,$$

where E_i and E_j are orthogonal if $i \neq j$, and thus, we must have $E_i = U_i$. Since $\mu_i, \mu_j \geq 0$ and $\mu_i \neq \mu_j$ implies that $\mu_i^2 \neq \mu_j^2$, the values $\mu_1^2, \ldots \mu_p^2$ are the distinct eigenvalues of f, and the corresponding eigenspaces are also $E_1, \ldots, E_p$. This shows that $g = h$, and h is unique. Also, as a consequence, $\mathrm{Ker}\, f = \mathrm{Ker}\, h$, and if $\mu_1, \ldots, \mu_p$ are the distinct eigenvalues of h, then $\mu_1^2, \ldots, \mu_p^2$ are the distinct eigenvalues of f, and the corresponding eigenspaces are identical. $\square$

There are now two ways to proceed. We can prove directly the singular value decomposition, as Strang does [25, 24], or prove the so-called *polar decomposition* theorem. The proofs are roughly of the same difficulty. We choose the second approach since it is less common in textbook presentations, and since it also yields a little more, namely uniqueness when f is invertible. It is somewhat disconcerting that the next two theorems are only given as an exercise in Bourbaki [6] (Algèbre, Chapter 9, problem 14, page 127). Yet, the SVD decomposition is of great practical importance. This is probably typical of the attitude of "pure mathematicians". However, the proof hinted at in Bourbaki is quite elegant. The early history of the Singular Value Decomposition is described in a fascinating paper by Stewart [23]. The SVD is due to Beltrami and Camille Jordan independently (1873, 1874). Gauss is the grand father of all this, for his work on least squares (1809, 1823) (but Legendre also published a paper on least squares!). Then come Sylvester, Schmidt, and Hermann Weyl. Sylvester's work was apparently "opaque". He gave some computational method to find an SVD. Schimdt's work really has to do with integral equations and symmetric and unsymmetric kernels (1907). Weyl's work has to due with perturbation theory (1912). Autonne came up with the polar decomposition (1902, 1915). Eckart and Young extended SVD to rectangular matrices (1936, 1939).

The next three theorems deal with a linear map $f\colon E \to E$ over a Euclidean space E. We will show later on how to generalize these results to linear maps $f\colon E \to F$ between two Euclidean spaces E and F.

Theorem 5.3 *Given a Euclidean space E of dimension n, for every linear map $f\colon E \to E$, there are two positive self-adjoint linear maps $h_1\colon E \to E$ and $h_2\colon E \to E$ and an orthogonal linear map $g\colon E \to E$ such that*

$$f = g \circ h_1 = h_2 \circ g.$$

Furthermore, if f has rank r, the maps h_1 and h_2 have the same positive eigenvalues $\mu_1, \ldots, \mu_r$, which are the singular values of f, i.e., the positive square roots of the non-null eigenvalues of both $f^* \circ f$ and $f \circ f^*$. Finally, g, h_1, h_2 are unique if f is invertible, and $h_1 = h_2$ if f is normal.

Proof. By Lemma 5.2, there are two (unique) positive self-adjoint linear maps $h_1\colon E \to E$ and $h_2\colon E \to E$ such that $f^* \circ f = h_1^2$ and $f \circ f^* = h_2^2$. Note that

$$\langle f(\overrightarrow{u}), f(\overrightarrow{v})\rangle = \langle h_1(\overrightarrow{u}), h_1(\overrightarrow{v})\rangle$$

for all $\overrightarrow{u}, \overrightarrow{v} \in E$, since

$$\langle f(\overrightarrow{u}), f(\overrightarrow{v})\rangle = \langle \overrightarrow{u}, (f^* \circ f)(\overrightarrow{v})\rangle = \langle \overrightarrow{u}, (h_1 \circ h_1)(\overrightarrow{v})\rangle = \langle h_1(\overrightarrow{u}), h_1(\overrightarrow{v})\rangle,$$

because $f^* \circ f = h_1^2$ and $h_1 = h_1^*$ (h_1 is self-adjoint). From Lemma 5.1, $\mathrm{Ker}\, f = \mathrm{Ker}\,(f^* \circ f)$, and from Lemma 5.2, $\mathrm{Ker}\,(f^* \circ f) = \mathrm{Ker}\, h_1$. Thus,

$$\mathrm{Ker}\, f = \mathrm{Ker}\, h_1.$$

If r is the rank of f, since h_1 is self-adjoint, by Theorem 3.1, there is an orthonormal basis $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ of eigenvectors of h_1, and by reordering these vectors if necessary, we can assume that $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_r})$ are associated with the strictly positive eigenvalues $\mu_1, \ldots, \mu_r$ of h_1 (the singular values of f), and that $\mu_{r+1} = \ldots = \mu_n = 0$. Observe that $(\overrightarrow{u_{r+1}}, \ldots, \overrightarrow{u_n})$ is an orthonormal basis of $\mathrm{Ker}\, f = \mathrm{Ker}\, h_1$, and that $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_r})$ is an orthonormal basis of $(\mathrm{Ker}\, f)^\perp = \mathrm{Im}\, f^*$. Note that

$$\langle f(\overrightarrow{u_i}), f(\overrightarrow{u_j})\rangle = \langle h_1(\overrightarrow{u_i}), h_1(\overrightarrow{u_j})\rangle = \mu_i\mu_j\langle \overrightarrow{u_i}, \overrightarrow{u_j}\rangle = \mu_i^2\delta_{i\,j}$$

when $1 \leq i, j \leq n$ (recall that $\delta_{i\,j} = 1$ iff $i = j$, and $\delta_{i\,j} = 0$ iff $i \neq j$). Letting

$$\overrightarrow{v_i} = \frac{f(\overrightarrow{u_i})}{\mu_i}$$

when $1 \leq i \leq r$, observe that

$$\langle \overrightarrow{v_i}, \overrightarrow{v_j}\rangle = \delta_{i\,j}$$

when $1 \leq i, j \leq r$. Using the Gram–Schmidt orthonormalization procedure, we can extend $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_r})$ to an orthonormal basis $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$ of E (even when $r = 0$). Also note that $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_r})$ is an orthonormal basis of $\mathrm{Im}\, f$, and $(\overrightarrow{v_{r+1}}, \ldots, \overrightarrow{v_n})$ is an orthonormal basis of $\mathrm{Im}\, f^{\perp} = \mathrm{Ker}\, f^*$.

We define the linear map $g\colon E \to E$ by its action on the basis $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ as follows:

$$g(\overrightarrow{u_i}) = \overrightarrow{v_i}$$

for all i, $1 \leq i \leq n$. We have

$$(g \circ h_1)(\overrightarrow{u_i}) = g(h_1(\overrightarrow{u_i})) = g(\mu_i \overrightarrow{u_i}) = \mu_i g(\overrightarrow{u_i}) = \mu_i \overrightarrow{v_i} = \mu_i \frac{f(\overrightarrow{u_i})}{\mu_i} = f(\overrightarrow{u_i})$$

when $1 \leq i \leq r$, and

$$(g \circ h_1)(\overrightarrow{u_i}) = g(h_1(\overrightarrow{u_i})) = g(\overrightarrow{0}) = \overrightarrow{0}$$

when $r+1 \leq i \leq n$ (since $(\overrightarrow{u_{r+1}}, \ldots, \overrightarrow{u_n})$ is a basis for $\mathrm{Ker}\, f = \mathrm{Ker}\, h_1$), which shows that $f = g \circ h_1$. The fact that g is orthogonal follows easily from the fact that it maps the orthonormal basis $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ to the orthonormal basis $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$.

We can show that $f = h_2 \circ g$ as follows. Notice that

$$\begin{aligned}
h_2^2(\overrightarrow{v_i}) &= (f \circ f^*)\left(\frac{f(\overrightarrow{u_i})}{\mu_i}\right), \\
&= (f \circ (f^* \circ f))\left(\frac{\overrightarrow{u_i}}{\mu_i}\right), \\
&= \frac{1}{\mu_i}(f \circ h_1^2)(\overrightarrow{u_i}), \\
&= \frac{1}{\mu_i} f(h_1^2(\overrightarrow{u_i})), \\
&= \frac{1}{\mu_i} f(\mu_i^2 \overrightarrow{u_i}), \\
&= \mu_i f(\overrightarrow{u_i}), \\
&= \mu_i^2 \overrightarrow{v_i}
\end{aligned}$$

when $1 \leq i \leq r$, and

$$h_2^2(\overrightarrow{v_i}) = (f \circ f^*)(\overrightarrow{v_i}) = f(f^*(\overrightarrow{v_i})) = \overrightarrow{0}$$

when $r+1 \leq i \leq n$, since $(\overrightarrow{v_{r+1}}, \ldots, \overrightarrow{v_n})$ is a basis for $\mathrm{Ker}\, f^* = (\mathrm{Im}\, f)^{\perp}$. Since h_2 is positive self-adjoint, so is h_2^2, and by Lemma 5.2, we must have

$$h_2(\overrightarrow{v_i}) = \mu_i \overrightarrow{v_i}$$

when $1 \leq i \leq r$, and

$$h_2(\overrightarrow{v_i}) = \overrightarrow{0}$$

when $r+1 \leq i \leq n$. This shows that $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$ are eigenvectors of h_2 for $\mu_1, \ldots, \mu_n$ (since $\mu_{r+1} = \ldots = \mu_n = 0$), and thus h_1 and h_2 have the same eigenvalues $\mu_1, \ldots, \mu_n$.

As a consequence,

$$(h_2 \circ g)(\overrightarrow{u_i}) = h_2(g(\overrightarrow{u_i})) = h_2(\overrightarrow{v_i}) = \mu_i \overrightarrow{v_i} = f(\overrightarrow{u_i})$$

when $1 \leq i \leq n$. Since h_1, h_2, $f^* \circ f$, and $f \circ f^*$ are positive self-adjoint, $f^* \circ f = h_1^2$, $f \circ f^* = h_2^2$, and $\mu_1, \ldots, \mu_r$ are the eigenvalues of both h_1 and h_2, it follows that $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e., the positive square roots of the nonnull eigenvalues of both $f^* \circ f$ and $f \circ f^*$.

Finally, if f is invertible, then h_1 and h_2 are invertible, and since

$$f^* \circ f = h_1^2 \quad \text{and} \quad f^* \circ f = h_1^2,$$

by Lemma 5.2, h_1 and h_2 are unique, and thus g is also unique since $g = f \circ h_1^{-1}$. If h is normal, $f^* \circ f = f \circ f^*$, and $h_1 = h_2$. $\square$

In matrix form, Theorem 5.4 can be stated as follows. For every real $n \times n$ matrix A, there is some orthogonal matrix R and some positive symmetric matrix S such that

$$A = RS.$$

Furthermore, R, S are unique if A is invertible. A pair (R, S) such that $A = RS$ is called a *polar decomposition of* A. For example, the matrix

$$A = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

is both orthogonal and symmetric, and $A = RS$ with $R = A$ and $S = I$, which implies that some of the eigenvalues of A are negative.

Remark: If E is a Hermitian space, Theorem 5.3 also holds, but the orthogonal linear map g becomes a unitary map. In terms of matrices, the polar decomposition states that for every complex $n \times n$ matrix A, there is some unitary matrix U and some positive Hermitian matrix H such that

$$A = UH.$$

The proof of Theorem 5.3 shows that there are two orthonormal bases $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$, where $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ are eigenvectors of h_1 and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$ are eigenvectors of h_2. Furthermore, $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_r})$ is an orthonormal basis of $\mathrm{Im}\, f^*$, $(\overrightarrow{u_{r+1}}, \ldots, \overrightarrow{u_n})$ is an orthonormal basis of $\mathrm{Ker}\, f$, $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_r})$ is an orthonormal basis of $\mathrm{Im}\, f$, and $(\overrightarrow{v_{r+1}}, \ldots, \overrightarrow{v_n})$ is an orthonormal basis of $\mathrm{Ker}\, f^*$. Using this, we immediately obtain the singular value decomposition theorem. I must say (I'm being a bit facetious) that I have come accross a textbook in which the singular value decomposition for linear maps of determinant $+1$ is called the *Cartan decomposition* (after Elie Cartan)!

Theorem 5.4 *Given a Euclidean space E of dimension n, for every linear map $f\colon E \to E$, there are two orthonormal bases $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$ such that if r is the rank of f, the matrix of f w.r.t. these two bases is a diagonal matrix of the form*

$$\begin{pmatrix} \mu_1 & & \ldots & \\ & \mu_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \mu_n \end{pmatrix}$$

where $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e. the (positive) square roots of the nonnull eigenvalues of $f^* \circ f$ and $f \circ f^*$, and $\mu_{r+1} = \ldots = \mu_n = 0$. Furthermore, $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ are eigenvectors of $f^* \circ f$, $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$ are eigenvectors of $f \circ f^*$, and $f(\overrightarrow{u_i}) = \mu_i \overrightarrow{v_i}$ when $1 \leq i \leq n$.

Proof. Going back to the proof of Theorem 5.4, there are two orthonormal bases $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$, where $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ are eigenvectors of h_1, $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$ are eigenvectors of h_2, and $f(\overrightarrow{u_i}) = \mu_i \overrightarrow{v_i}$ when $1 \leq i \leq r$, and $f(\overrightarrow{u_i}) = \overrightarrow{0}$ when $r+1 \leq i \leq n$. But now, with respect to the orthonormal bases $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$, the matrix of f is indeed

$$\begin{pmatrix} \mu_1 & & \ldots & \\ & \mu_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \mu_n \end{pmatrix}$$

where $\mu_1, \ldots, \mu_r$ are the singular values of f and $\mu_{r+1} = \ldots = \mu_n = 0$. $\square$

Note that $\mu_i > 0$ for all i ($1 \leq i \leq n$) iff f is invertible. Given an orientation of the Euclidean space E specified by some orthonormal basis $(\overrightarrow{e_1}, \ldots, \overrightarrow{e_n})$ taken as direct, if $\det(f) \geq 0$, we can always make sure that the two orthonormal bases $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$ are oriented positively. Indeed, if $\det(f) = 0$, we just have to flip $\overrightarrow{u_n}$ to $-\overrightarrow{u_n}$ if necessary, and $\overrightarrow{v_n}$ to $-\overrightarrow{v_n}$ if necessary. If $\det(f) > 0$, since $\mu_i > 0$ for all i, $1 \leq i \leq n$, the orthogonal matrices U and V whose columns are the $\overrightarrow{u_i}$'s and the $\overrightarrow{v_i}$'s have determinants of the same sign. Since $f(\overrightarrow{u_n}) = \mu_n \overrightarrow{v_n}$ and $\mu_n > 0$, we just have to flip $\overrightarrow{u_n}$ to $-\overrightarrow{u_n}$ if necessary, since $\overrightarrow{v_n}$ will also be flipped. Theorem 5.4 can be restated in terms of (real) matrices as follows.

Theorem 5.5 *For every real $n \times n$ matrix A, there are two orthogonal matrices U and V and a diagonal matrix D such that $A = V D U^\top$, where D is of the form*

$$D = \begin{pmatrix} \mu_1 & & \ldots & \\ & \mu_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \mu_n \end{pmatrix}$$

where $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e. the (positive) square roots of the nonnull eigenvalues of $A^\top A$ and $A\,A^\top$, and $\mu_{r+1} = \ldots = \mu_n = 0$. The columns of U are eigenvectors of $A^\top A$, and the columns of V are eigenvectors of $A\,A^\top$. Furthermore, if $\det(A) \geq 0$, it is possible to choose U and V so that $\det(U) = \det(V) = +1$, i.e., U and V are rotation matrices.

A triple (U, D, V) such that $A = V D U^\top$ is called a *singular value decomposition (SVD) of A*.

Remark: In Strang [25], the matrices U, V, D are denoted as $U = Q_2$, $V = Q_1$, and $D = \Sigma$, and a SVD decomposition is written as $A = Q_1 D Q_2^\top$. This has the advantage that Q_1 comes before Q_2 in $A = Q_1 D Q_2^\top$. This has the disadvantage that A maps the columns of Q_2 (eigenvectors of $A^\top A$) to multiples of the columns of Q_1 (eigenvectors of $A\,A^\top$).

The SVD also applies to complex matrices. In this case, for every complex $n \times n$ matrix A, there are two unitary matrices U and V and a diagonal matrix D such that

$$A = V D U^*,$$

where D is a diagonal matrix consisting of real entries $\mu_1, \ldots, \mu_n$, where $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e. the positive square roots of the nonnull eigenvalues of $A^* A$ and $A\,A^*$, and $\mu_{r+1} = \ldots = \mu_n = 0$.

It is easy to go from the polar form to the SVD, and backward. Indeed, given a polar decomposition $A = R_1 S$, where R_1 is orthogonal and S is positive symmetric, there is an orthogonal matrix R_2 and a positive diagonal matrix D such that $S = R_2 D\, R_2^\top$, and thus

$$A = R_1 R_2 D\, R_2^\top = V D U^\top,$$

where $V = R_1 R_2$ and $U = R_2$ are orthogonal.

Going the other way, given an SVD decomposition $A = VDU^\top$, let $R = VU^\top$ and $S = UDU^\top$. It is clear that R is orthogonal and that S is positive symmetric, and

$$RS = VU^\top UDU^\top = VDU^\top = A.$$

Note that it is possible to require that $\det(R) = +1$ when $\det(A) \geq 0$.

Theorem 5.5 can be easily extended to rectangular $m \times n$ matrices (see Strang [25]). As a matter of fact, both Theorems 5.3 and 5.4 can be generalized to linear maps $f: E \to F$ between two Euclidean spaces E and F. In order to do so, we need to define the analog of the notion of an orthogonal linear map for linear maps $f: E \to F$. By definition, the adjoint $f^*: F \to E$ of a linear map $f: E \to F$ is the unique linear map such that

$$\langle f(\overrightarrow{u}), \overrightarrow{v} \rangle_2 = \langle \overrightarrow{u}, f^*(\overrightarrow{v}) \rangle_1$$

for all $\overrightarrow{u} \in E$ and all $\overrightarrow{v} \in F$. Then, we have

$$\langle f(\overrightarrow{u}), f(\overrightarrow{v}) \rangle_2 = \langle \overrightarrow{u}, (f^* \circ f)(\overrightarrow{v}) \rangle_1$$

for all $\overrightarrow{u}, \overrightarrow{v} \in E$. Letting $n = \dim(E)$, $m = \dim(F)$, and $p = \min(m, n)$, if f has rank p and if for every p orthonormal vectors $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_p})$ in $(\operatorname{Ker} f)^\perp$, the vectors $(f(\overrightarrow{u_1}), \ldots, f(\overrightarrow{u_p}))$ are also orthonormal in F, then

$$f^* \circ f = \mathrm{id}$$

on $(\operatorname{Ker} f)^\perp$. The converse is immediately proved. Thus, we will say that a linear map $f: E \to F$ is *weakly orthogonal* if is has rank $p = \min(m, n)$ and if

$$f^* \circ f = \mathrm{id}$$

on $(\operatorname{Ker} f)^\perp$. Of course, $f^* \circ f = 0$ on $\operatorname{Ker} f$. In terms of matrices, we will say that a real $m \times n$ matrix A is weakly orthogonal if its first $p = \min(m, n)$ columns are orthonormal, the remaining ones (if any) being null columns. This is equivalent to saying that

$$A^\top A = I_n$$

if $m \geq n$, and that

$$A^\top A = \begin{pmatrix} I_m & 0_{m,n-m} \\ 0_{n-m,m} & 0_{n-m,n-m} \end{pmatrix}$$

if $n > m$. In this latter case $(n > m)$, it is immediately shown that

$$A\,A^\top = I_m,$$

and $A^\top$ is also weakly orthogonal. The main difference with orthogonal matrices is that $A\,A^\top$ is usually not a nice matrix of the above form when $m \geq n$ (unless $m = n$). Weakly unitary linear maps are defined analogously.

Theorem 5.6 *Given any two Euclidean spaces E and F, where E has dimension n and F has dimension m, for every linear map $f\colon E \to F$, there are two positive self-adjoint linear maps $h_1\colon E \to E$ and $h_2\colon F \to F$ and a weakly orthogonal linear map $g\colon E \to F$ such that*

$$f = g \circ h_1 = h_2 \circ g.$$

Furthermore, if f has rank r, the maps h_1 and h_2 have the same positive eigenvalues $\mu_1, \ldots, \mu_r$, which are the singular values of f, i.e., the positive square roots of the non-null eigenvalues of both $f^* \circ f$ and $f \circ f^*$. Finally, g, h_1, h_2 are unique if f is invertible, and $h_1 = h_2$ if f is normal.

Proof. By Lemma 5.2, there are two (unique) positive self-adjoint linear maps $h_1\colon E \to E$ and $h_2\colon F \to F$ such that $f^* \circ f = h_1^2$ and $f \circ f^* = h_2^2$. As in the proof of Theorem 5.3,

$$\operatorname{Ker} f = \operatorname{Ker} h_1,$$

and letting r be the rank of f, there is an orthonormal basis $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ of eigenvectors of h_1 such that $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_r})$ are associated with the strictly positive eigenvalues $\mu_1, \ldots, \mu_r$ of h_1 (the singular values of f). The vectors $(\overrightarrow{u_{r+1}}, \ldots, \overrightarrow{u_n})$ form an orthonormal basis of $\operatorname{Ker} f = \operatorname{Ker} h_1$, and the vectors $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_r})$ form an orthonormal basis of $(\operatorname{Ker} f)^\perp = \operatorname{Im} f^*$. Letting

$$\overrightarrow{v_i} = \frac{f(\overrightarrow{u_i})}{\mu_i}$$

when $1 \leq i \leq r$, using the Gram–Schmidt orthonormalization procedure, we can extend $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_r})$ to an orthonormal basis $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$ of F (even when $r = 0$). Also note that $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_r})$ is an orthonormal basis of $\operatorname{Im} f$, and $(\overrightarrow{v_{r+1}}, \ldots, \overrightarrow{v_m})$ is an orthonormal basis of $\operatorname{Im} f^\perp = \operatorname{Ker} f^*$.

Letting $p = \min(m, n)$, we define the linear map $g\colon E \to F$ by its action on the basis $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ as follows:

$$g(\overrightarrow{u_i}) = \overrightarrow{v_i}$$

for all i, $1 \leq i \leq p$, and

$$g(\overrightarrow{u_i}) = \overrightarrow{0}$$

for all i, $p + 1 \leq i \leq n$. Note that $r \leq p$. Just as in the proof of Theorem 5.3, we have

$$(g \circ h_1)(\overrightarrow{u_i}) = f(\overrightarrow{u_i})$$

when $1 \leq i \leq r$, and

$$(g \circ h_1)(\overrightarrow{u_i}) = g(h_1(\overrightarrow{u_i})) = g(\overrightarrow{0}) = \overrightarrow{0}$$

when $r + 1 \leq i \leq n$ (since $(\overrightarrow{u_{r+1}}, \ldots, \overrightarrow{u_n})$ is a basis for $\operatorname{Ker} f = \operatorname{Ker} h_1$), which shows that $f = g \circ h_1$. The fact that g is weakly orthogonal follows easily from the fact that it maps the orthonormal vectors $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_p})$ to the orthonormal vectors $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_p})$.

We can show that $f = h_2 \circ g$ as follows. Just as in the proof of Theorem 5.3,

$$h_2^2(\overrightarrow{v_i}) = \mu_i^2 \overrightarrow{v_i}$$

when $1 \leq i \leq r$, and

$$h_2^2(\overrightarrow{v_i}) = (f \circ f^*)(\overrightarrow{v_i}) = f(f^*(\overrightarrow{v_i})) = \overrightarrow{0}$$

when $r+1 \leq i \leq m$, since $(\overrightarrow{v_{r+1}}, \ldots, \overrightarrow{v_m})$ is a basis for $\mathrm{Ker}\, f^* = (\mathrm{Im}\, f)^{\perp}$. Since h_2 is positive self-adjoint, so is h_2^2, and by Lemma 5.2, we must have

$$h_2(\overrightarrow{v_i}) = \mu_i \overrightarrow{v_i}$$

when $1 \leq i \leq r$, and

$$h_2(\overrightarrow{v_i}) = \overrightarrow{0}$$

when $r+1 \leq i \leq m$. This shows that $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$ are eigenvectors of h_2 for $\mu_1, \ldots, \mu_m$ (letting $\mu_{r+1} = \ldots = \mu_m = 0$), and thus h_1 and h_2 have the same nonnull eigenvalues $\mu_1, \ldots, \mu_r$.

As a consequence,

$$(h_2 \circ g)(\overrightarrow{u_i}) = h_2(g(\overrightarrow{u_i})) = h_2(\overrightarrow{v_i}) = \mu_i \overrightarrow{v_i} = f(\overrightarrow{u_i})$$

when $1 \leq i \leq m$. Since h_1, h_2, $f^* \circ f$, and $f \circ f^*$ are positive self-adjoint, $f^* \circ f = h_1^2$, $f \circ f^* = h_2^2$, and $\mu_1, \ldots, \mu_r$ are the eigenvalues of both h_1 and h_2, it follows that $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e., the positive square roots of the nonnull eigenvalues of both $f^* \circ f$ and $f \circ f^*$.

Finally, if f is invertible, then h_1 and h_2 are invertible, and since

$$f^* \circ f = h_1^2 \quad \text{and} \quad f^* \circ f = h_1^2,$$

by Lemma 5.2, h_1 and h_2 are unique, and thus g is also unique since $g = f \circ h_1^{-1}$. If h is normal, $f^* \circ f = f \circ f^*$, and $h_1 = h_2$. □

In matrix form, Theorem 5.6 can be stated as follows. For every real $m \times n$ matrix A, there is some weakly orthogonal $m \times n$ matrix R and some positive symmetric $n \times n$ matrix S such that

$$A = RS.$$

The proof also shows that if $n > m$, the last $n - m$ columns of R are zero vectors. A pair (R, S) such that $A = RS$ is called a *polar decomposition of* A.

Remark: If E is a Hermitian space, Theorem 5.6 also holds, but the weakly orthogonal linear map g becomes a weakly unitary map. In terms of matrices, the polar decomposition states that for every complex $m \times n$ matrix A, there is some weakly unitary $m \times n$ matrix U and some positive Hermitian $n \times n$ matrix H such that

$$A = UH.$$

The proof of Theorem 5.6 shows that there are two orthonormal bases $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ of E and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$ of F, where $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ are eigenvectors of h_1 and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$ are eigenvectors of h_2. Furthermore, $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_r})$ is an orthonormal basis of $\operatorname{Im} f^*$, $(\overrightarrow{u_{r+1}}, \ldots, \overrightarrow{u_n})$ is an orthonormal basis of $\operatorname{Ker} f$, $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_r})$ is an orthonormal basis of $\operatorname{Im} f$, and $(\overrightarrow{v_{r+1}}, \ldots, \overrightarrow{v_m})$ is an orthonormal basis of $\operatorname{Ker} f^*$. Using this, we immediately obtain the singular value decomposition theorem for linear maps $f \colon E \to F$, where E and F can have different dimensions.

Theorem 5.7 *Given any two Euclidean spaces E and F, where E has dimension n and F has dimension m, for every linear map $f \colon E \to F$, there are two orthonormal bases $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ and $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$ such that if r is the rank of f, the matrix of f w.r.t. these two bases is a $m \times n$ matrix D of the form*

$$D = \begin{pmatrix} \mu_1 & & \ldots & \\ & \mu_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \mu_n \\ 0 & \vdots & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \ldots & 0 \end{pmatrix} \quad or \quad D = \begin{pmatrix} \mu_1 & & \ldots & & 0 & \ldots & 0 \\ & \mu_2 & \ldots & & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \vdots & 0 \\ & & \ldots & \mu_m & 0 & \ldots & 0 \end{pmatrix}$$

where $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e. the (positive) square roots of the nonnull eigenvalues of $f^* \circ f$ and $f \circ f^*$, and $\mu_{r+1} = \ldots = \mu_p = 0$, where $p = \min(m, n)$. Furthermore, $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ are eigenvectors of $f^* \circ f$, $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_m})$ are eigenvectors of $f \circ f^*$, and $f(\overrightarrow{u_i}) = \mu_i \overrightarrow{v_i}$ when $1 \leq i \leq p = \min(m, n)$.

Even though the matrix D is an $m \times n$ rectangular matrix, since its only nonzero entries are on the descending diagonal, we still say that D is a diagonal matrix. Theorem 5.7 can be restated in terms of (real) matrices as follows.

Theorem 5.8 *For every real $m \times n$ matrix A, there are two orthogonal matrices U $(n \times n)$ and V $(m \times m)$ and a diagonal $m \times n$ matrix D such that $A = V D U^\top$, where D is of the form*

$$D = \begin{pmatrix} \mu_1 & & \ldots & \\ & \mu_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \mu_n \\ 0 & \vdots & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \ldots & 0 \end{pmatrix} \quad or \quad D = \begin{pmatrix} \mu_1 & & \ldots & & 0 & \ldots & 0 \\ & \mu_2 & \ldots & & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \vdots & 0 \\ & & \ldots & \mu_m & 0 & \ldots & 0 \end{pmatrix}$$

where $\mu_1, \ldots, \mu_r$ are the singular values of f, i.e. the (positive) square roots of the nonnull eigenvalues of $A^\top A$ and $A\,A^\top$, and $\mu_{r+1} = \ldots = \mu_p = 0$, where $p = \min(m, n)$. The columns of U are eigenvectors of $A^\top A$, and the columns of V are eigenvectors of $A\,A^\top$.

A triple (U, D, V) such that $A = V\,D\,U^\top$ is called a *singular value decomposition (SVD) of A*. The SVD decomposition of matrices can be used to define the pseudo-inverse of a rectangular matrix, see Strang [25] for a thorough presentation.

Remark: The matrix form of Theorem 5.6 also yields a variant of the singular value decomposition. First, assume that $m \geq n$. Given an $m \times n$ matrix A, there is a weakly orthogonal $m \times n$ matrix R_1 and a positive symmetric $n \times n$ matrix S, such that

$$A = R_1 S.$$

Since S is positive symmetric, there is an orthogonal $n \times n$ matrix R_2 and a diagonal $n \times n$ matrix D with nonnegative entries, such that

$$S = R_2 D\, R_2^\top.$$

Thus, we can write

$$A = R_1 R_2 D\, R_2^\top.$$

We claim that $R_1 R_2$ is weakly orthogonal. Indeed,

$$(R_1 R_2)^\top (R_1 R_2) = R_2^\top\, (R_1^\top R_1) R_2,$$

and if $m \geq n$, we have

$$R_1^\top R_1 = I_n$$

so that

$$(R_1 R_2)^\top (R_1 R_2) = I_n.$$

Thus, $R_1 R_2$ is indeed weakly orthogonal. Let us now consider the case $n > m$. From the version of SVD in which

$$A = V\,D\,U^\top$$

where U is $n \times n$ orthogonal, V is $m \times m$ orthogonal, and D is $m \times n$ diagonal with nonnegative diagonal entries, letting V' be the $m \times n$ matrix obtained from V by adding $n - m$ zero columns and D' be the $n \times n$ matrix obtained from D by adding $n - m$ zero rows, it is immediately verified that

$$V' D' = V D,$$

and thus, when $n > m$, we also have

$$A = V' D'\, U^\top,$$

where U is $n \times n$ orthogonal, V' is $m \times n$ weakly orthogonal, and D' is $n \times n$ diagonal with nonnegative diagonal entries. As a consequence, in both cases, we have shown that there exists a weakly orthogonal $m \times n$ matrix V, an orthogonal $n \times n$ matrix U, and a diagonal $n \times n$ matrix D with nonnegative entries, such that

$$A = V D U^\top.$$

There is yet another alternative when $n > m$. Given an $m \times n$ matrix A, there is a positive symmetric $m \times m$ matrix S and a weakly orthogonal $m \times n$ matrix R_1, such that

$$A = S R_1.$$

Since S is positive symmetric, there is an orthogonal $m \times m$ matrix R_2 an a diagonal $m \times m$ matrix D with nonnegative entries, such that

$$S = R_2 D\, R_2^\top.$$

Thus, we can write

$$A = R_2 D\, R_2^\top R_1.$$

We claim that $R_2^\top R_1$ is weakly orthogonal. Indeed,

$$(R_2^\top R_1)^\top R_2^\top R_1 = R_1^\top (R_2 R_2^\top) R_1 = R_1^\top R_1,$$

since R_2 is orthogonal, and if $n > m$, we have

$$R_1^\top R_1 = \begin{pmatrix} I_m & 0_{m,n-m} \\ 0_{n-m,m} & 0_{n-m,n-m} \end{pmatrix},$$

so that

$$(R_2^\top R_1)^\top R_2^\top R_1 = \begin{pmatrix} I_m & 0_{m,n-m} \\ 0_{n-m,m} & 0_{n-m,n-m} \end{pmatrix},$$

and $R_2^\top R_1$ is weakly orthogonal. Since $n > m$, $(R_2^\top R_1)^\top = R_1^\top R_2$ is also weakly orthogonal. As a consequence, we have shown that when $m \geq n$, there exists a weakly orthogonal $m \times n$ matrix V, an orthogonal $n \times n$ matrix U, and a diagonal $n \times n$ matrix D with nonnegative entries, such that

$$A = V D U^\top,$$

and when $n > m$, there exists an orthogonal $m \times m$ matrix V, a weakly orthogonal $m \times n$ matrix $U^\top$ (with U also weakly orthogonal), and a diagonal $m \times m$ matrix D with nonnegative entries, such that

$$A = V D U^\top.$$

In both cases,

$$V^\top A U = D.$$

6 The Exponential Map

In this section, we introduce the exponential map on matrices and show some of its properties. The exponential map is a very valuable tool that allows us to "linearize" certain algebraic properties of matrices. It also plays a crucial role in the theory of linear differential equations with constant coefficients. But most of all (at least for us), it is a stepping stone to Lie groups and Lie algebras. This section is inspired from Artin [2], Chevalley [9], Marsden and Ratiu [17], Curtis [10], Howe [14], and Sattinger and Weaver [20].

Given an $n \times n$ (real or complex) matrix $A = (a_{i,j})$, we would like to define the exponential e^A of A as the sum of the series

$$e^A = I_n + \sum_{p \geq 1} \frac{A^p}{p!} = \sum_{p \geq 0} \frac{A^p}{p!},$$

letting $A^0 = I_n$. The problem is, why is it well defined? The following Lemma shows that the above series is indeed absolutely convergent.

Lemma 6.1 *Let $A = (a_{i\,j})$ be a (real or complex) $n \times n$ matrix, and let*

$$\mu = \max\{|a_{i\,j}| \mid 1 \leq i, j \leq n\}.$$

If $A^p = (a^p_{i\,j})$, then

$$|a^p_{i\,j}\,| \leq (n\mu)^p$$

for all i, j, $1 \leq i, j \leq n$. As a consequence, the n^2 series

$$\sum_{p \geq 0} \frac{a^p_{i\,j}}{p!}$$

converge absolutely, and the matrix

$$e^A = \sum_{p \geq 0} \frac{A^p}{p!}$$

is a well defined matrix.

Proof. The proof is by induction on p. For $p = 0$, $A^0 = I_n$, $(n\mu)^0 = 1$, and the lemma is obvious. Assume that

$$|a^p_{i\,j}| \leq (n\mu)^p$$

for all i, j, $1 \leq i, j \leq n$. Then, we have

$$|a^{p+1}_{i\,j}| = |\sum_{k=1}^{n} a^p_{i\,k} a_{k\,j}| \leq \sum_{k=1}^{n} |a^p_{i\,k}||a_{k\,j}| \leq \mu \sum_{k=1}^{n} |a^p_{i\,k}| \leq n\mu(n\mu)^p = (n\mu)^{p+1},$$

for all i, j, $1 \leq i, j \leq n$. For every pair (i, j) such that $1 \leq i, j \leq n$, since

$$|a^p_{i\,j}| \leq (n\mu)^p,$$

the series

$$\sum_{p \geq 0} \frac{|a^p_{i\,j}|}{p!}$$

is bounded by the convergent series

$$e^{n\mu} = \sum_{p \geq 0} \frac{(n\mu)^p}{p!},$$

and thus, it is absoluteley convergent. This shows that

$$e^A = \sum_{k \geq 0} \frac{A^k}{k!}$$

is well defined. $\square$

It is instructive to compute explicitly the exponential of some simple matrices. As an example, let us compute the exponential of the real skew symmetric matrix

$$A = \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}.$$

We need to find an inductive formula expressing the powers A^n. Let us observe that

$$\begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix} = \theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}^2 = -\theta^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then, letting

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

we have

$$\begin{aligned} A^{4n} &= \theta^{4n} I_2, \\ A^{4n+1} &= \theta^{4n+1} J, \\ A^{4n+2} &= -\theta^{4n+2} I_2, \\ A^{4n+3} &= -\theta^{4n+3} J, \end{aligned}$$

and so

$$e^A = I_2 + \frac{\theta}{1!} J - \frac{\theta^2}{2!} I_2 - \frac{\theta^3}{3!} J + \frac{\theta^4}{4!} I_2 + \frac{\theta^5}{5!} J - \frac{\theta^6}{6!} I_2 - \frac{\theta^7}{7!} J + \cdots.$$

Rearranging the order of the terms, we have

$$e^A = \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots\right) I_2 + \left(\frac{\theta}{1!} - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots\right) J.$$

We recognize the power series for $\cos\theta$ and $\sin\theta$, and thus,

$$e^A = \cos\theta I_2 + \sin\theta J,$$

that is

$$e^A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Thus, e^A is a rotation matrix! This is a general fact. If A is a skew symmetric matrix, then e^A is an orthogonal matrix of determinant $+1$, i.e. a rotation matrix. Furthermore, every rotation matrix is of this form, i.e., the exponential map from the set of skew symmetric matrices to the set of rotation matrices is surjective. In order to prove these facts, we need to establish some properties of the exponential map. But before that, let us work out another example showing that the exponential map is not always surjective. Let us compute the exponential of a real 2×2 matrix with null trace of the form

$$A = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}.$$

We need to find an inductive formula expressing the powers A^n. Observe that

$$A^2 = (a^2 + bc)I_2 = -\det(A)I_2.$$

If $a^2 + bc = 0$, we have

$$e^A = I_2 + A.$$

If $a^2 + bc < 0$, let $\omega > 0$ be such that $\omega^2 = -(a^2 + bc)$. Then, $A^2 = -\omega^2 I_2$. We get

$$e^A = I_2 + \frac{A}{1!} - \frac{\omega^2}{2!}I_2 - \frac{\omega^2}{3!}A + \frac{\omega^4}{4!}I_2 + \frac{\omega^4}{5!}A - \frac{\omega^6}{6!}I_2 - \frac{\omega^6}{7!}A + \cdots.$$

Rearranging the order of the terms, we have

$$e^A = \left(1 - \frac{\omega^2}{2!} + \frac{\omega^4}{4!} - \frac{\omega^6}{6!} + \cdots\right) I_2 + \frac{1}{\omega}\left(\omega - \frac{\omega^3}{3!} + \frac{\omega^5}{5!} - \frac{\omega^7}{7!} + \cdots\right) A.$$

We recognize the power series for $\cos\omega$ and $\sin\omega$, and thus,

$$e^A = \cos\omega\, I_2 + \frac{\sin\omega}{\omega}A.$$

If $a^2 + bc > 0$, let $\omega > 0$ be such that $\omega^2 = (a^2 + bc)$. Then, $A^2 = \omega^2 I_2$. We get

$$e^A = I_2 + \frac{A}{1!} + \frac{\omega^2}{2!} I_2 + \frac{\omega^2}{3!} A + \frac{\omega^4}{4!} I_2 + \frac{\omega^4}{5!} A + \frac{\omega^6}{6!} I_2 + \frac{\omega^6}{7!} A + \cdots.$$

Rearranging the order of the terms, we have

$$e^A = \left(1 + \frac{\omega^2}{2!} + \frac{\omega^4}{4!} + \frac{\omega^6}{6!} + \cdots\right) I_2 + \frac{1}{\omega}\left(\omega + \frac{\omega^3}{3!} + \frac{\omega^5}{5!} + \frac{\omega^7}{7!} + \cdots\right) A.$$

If we recall that $\cosh \omega = \frac{e^{\omega} + e^{-\omega}}{2}$ and $\sinh \omega = \frac{e^{\omega} - e^{-\omega}}{2}$, we recognize the power series for $\cosh \omega$ and $\sinh \omega$, and thus,

$$e^A = \cosh \omega \, I_2 + \frac{\sinh \omega}{\omega} A.$$

It immediately verified that in all cases,

$$\det(e^A) = 1.$$

This shows that the exponential map is a function from the set of 2×2 matrices with null trace to the set of 2×2 matrices with determinant 1. This function is not surjective. Indeed, $tr(e^A) = 2\cos \omega$ when $a^2 + bc < 0$, $tr(e^A) = 2\cosh \omega$ when $a^2 + bc > 0$, and $tr(e^A) = 2$ when $a^2 + bc = 0$. As a consequence, for any matrix A with null trace,

$$tr(e^A) \geq -2,$$

and any matrix B with determinant 1 and whose trace is less than -2 is not the exponential e^A of any matrix A with null trace. For example,

$$B = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$$

where $a < 0$ and $a \neq -1$, is not the exponential of any matrix A with null trace.

A fundamental property of the exponential map is that if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, then the eigenvalues of e^A are $e^{\lambda_1}, \ldots, e^{\lambda_n}$. For this, we need two Lemmas.

Lemma 6.2 *Let A and U be (real or complex) matrices, and assume that U is invertible. Then,*

$$e^{UAU^{-1}} = Ue^AU^{-1}.$$

Proof. A trivial induction shows that

$$UA^pU^{-1} = (UAU^{-1})^p,$$

and thus,

$$e^{UAU^{-1}} = \sum_{p\geq 0} \frac{(UAU^{-1})^p}{p!} = \sum_{p\geq 0} \frac{UA^pU^{-1}}{p!} = U\left(\sum_{p\geq 0} \frac{A^p}{p!}\right)U^{-1} = Ue^AU^{-1}.$$

$\square$

Say that a square matrix A is an *upper triangular matrix* if it has the following shape

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & a_{1\,3} & \dots & a_{1\,n-1} & a_{1\,n} \\ 0 & a_{2\,2} & a_{2\,3} & \dots & a_{2\,n-1} & a_{2\,n} \\ 0 & 0 & a_{3\,3} & \dots & a_{3\,n-1} & a_{3\,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{n-1\,n-1} & a_{n-1\,n} \\ 0 & 0 & 0 & \dots & 0 & a_{n\,n} \end{pmatrix}$$

i.e., $a_{i\,j} = 0$ whenever $j < i$, $1 \leq i, j \leq n$.

Lemma 6.3 *Given any complex $n \times n$ matrix A, there is an invertible matrix P and an upper triangular matrix T such that*

$$A = PTP^{-1}.$$

Proof. We prove by induction on n that if $f\colon \mathbb{C}^n \to \mathbb{C}^n$ is a linear map, then there is a basis $(\overrightarrow{u_1}, \dots, \overrightarrow{u_n})$ with respect to which f is represented by an upper triangular matrix. For $n = 1$, the result is obvious. If $n > 1$, since $\mathbb{C}$ is algebraically closed, f has some eigenvalue $\lambda_1 \in \mathbb{C}$, and let $\overrightarrow{u_1}$ be an eigenvector for λ_1. We can find $n-1$ vectors $(\overrightarrow{v_2}, \dots, \overrightarrow{v_n})$ such that $(\overrightarrow{u_1}, \overrightarrow{v_2}, \dots, \overrightarrow{v_n})$ is a basis of $\mathbb{C}^n$, and let W be the subspace of dimension $n-1$ spanned by $(\overrightarrow{v_2}, \dots, \overrightarrow{v_n})$. In the basis $(\overrightarrow{u_1}, \overrightarrow{v_2} \dots, \overrightarrow{v_n})$, the matrix of f is the form

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & \dots & a_{1\,n} \\ 0 & a_{2\,2} & \dots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n\,2} & \dots & a_{n\,n} \end{pmatrix}$$

since its first column contains the coordinates of $\lambda_1\overrightarrow{u_1}$ over the basis $(\overrightarrow{u_1}, \overrightarrow{v_2}, \dots, \overrightarrow{v_n})$. Letting $p\colon \mathbb{C}^n \to W$ be the projection defined such that $p(\overrightarrow{u_1}) = \overrightarrow{0}$ and $p(\overrightarrow{v_i}) = \overrightarrow{v_i}$ when $2 \leq i \leq n$, the linear map $g\colon W \to W$ defined as the restriction of $p \circ f$ to W is represented by the $(n-1)\times(n-1)$ matrix $(a_{i\,j})_{2\leq i,j\leq n}$ over the basis $(\overrightarrow{v_2}, \dots, \overrightarrow{v_n})$. By the induction hypothesis, there is a basis $(\overrightarrow{u_2}, \dots, \overrightarrow{u_n})$ of W such that g is represented by an upper triangular matrix $(b_{i\,j})_{1\leq i,j\leq n-1}$.

However,

$$\mathbb{C}^n = \mathbb{C}\overrightarrow{u_1} \oplus W,$$

and thus, $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ is a basis for $\mathbb{C}^n$. Since p is the projection from $\mathbb{C}^n = \mathbb{C}\overrightarrow{u_1} \oplus W$ onto W and $g\colon W \to W$ is the restriction of $p \circ f$ to W, we have

$$f(\overrightarrow{u_1}) = \lambda_1 \overrightarrow{u_1}$$

and

$$f(\overrightarrow{u_{i+1}}) = a_{1\,i}\overrightarrow{u_1} + \sum_{j=1}^{n-1} b_{i\,j}\overrightarrow{u_{j+1}}$$

for some $a_{1\,i} \in \mathbb{C}$, when $1 \leq i \leq n-1$. But then, the matrix of f with respect to $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ is upper triangular. Thus, there is a change of basis matrix P such that $A = PTP^{-1}$ where T is upper triangular. $\square$

Remark: If E is a Hermitian space, the proof of Lemma 6.3 can be easily adapted to prove that there is an *orthonormal* basis $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ with respect to which the matrix of f is upper triangular. In terms of matrices, this means that there is a unitary matrix U and an upper triangular matrix T such that $A = UTU^*$. This is usually known as *Schur's Lemma*. Using this result, we can immediately rederive the fact that if A is a Hermitian matrix, then there is a unitary matrix U and a real diagonal matrix D such that $A = UDU^*$.

If $A = PTP^{-1}$ where T is upper triangular, note that the diagonal entries on T are the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A. Indeed, A and T have the same characteristic polynomial. This is because if A and B are any two matrices such that $A = PBP^{-1}$,

$$\begin{aligned}
\det(A - \lambda\, I) &= \det(PBP^{-1} - \lambda\, P\,I P^{-1}),\\
&= \det(P(B - \lambda\, I)P^{-1}),\\
&= \det(P)\det(B - \lambda\, I)\det(P^{-1}),\\
&= \det(P)\det(B - \lambda\, I)\det(P)^{-1},\\
&= \det(B - \lambda\, I).
\end{aligned}$$

Furthermore, it is well known that the determinant of a matrix of the form

$$\begin{pmatrix}
\lambda_1 - \lambda & a_{1\,2} & a_{1\,3} & \ldots & a_{1\,n-1} & a_{1\,n}\\
0 & \lambda_2 - \lambda & a_{2\,3} & \ldots & a_{2\,n-1} & a_{2\,n}\\
0 & 0 & \lambda_3 - \lambda & \ldots & a_{3\,n-1} & a_{3\,n}\\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\
0 & 0 & 0 & \ldots & \lambda_{n-1} - \lambda & a_{n-1\,n}\\
0 & 0 & 0 & \ldots & 0 & \lambda_n - \lambda
\end{pmatrix}$$

is $(\lambda_1 - \lambda)\cdots(\lambda_n - \lambda)$, and thus, the eigenvalues of $A = PTP^{-1}$ are the diagonal entries of T. We use this property to prove the following Lemma.

Lemma 6.4 *Given any complex $n \times n$ matrix A, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, then $e^{\lambda_1}, \ldots, e^{\lambda_n}$ are the eigenvalues of e^A. Furthermore, if $\overrightarrow{u}$ is an eigenvector of A for λ_i, then $\overrightarrow{u}$ is an eigenvector of e^A for e^{λ_i}.*

Proof. By lemma 6.3, there is an invertible matrix P and an upper diagonal matrix T such that

$$A = PTP^{-1}.$$

By Lemma 6.2,

$$e^{PTP^{-1}} = Pe^T P^{-1}.$$

However, we showed that A and T have the same eigenvalues, which are the diagonal entries $\lambda_1, \ldots, \lambda_n$ of T, and $e^A = e^{PTP^{-1}} = Pe^T P^{-1}$ and e^T have the same eigenvalues, which are the diagonal entries of e^T. Clearly, the diagonal entries of e^T are $e^{\lambda_1}, \ldots, e^{\lambda_n}$. Now, if $\overrightarrow{u}$ is an eigenvector of A for the eigenvalue λ, a simple induction shows that $\overrightarrow{u}$ is an eigenvector of A^n for the eigenvalue λ^n, from which is follows that $\overrightarrow{u}$ is an eigenvector of e^A for e^λ. $\square$

As a consequence, we can show that

$$\det(e^A) = e^{tr(A)},$$

where $tr(A)$ is the *trace of* A, i.e. the sum $a_{1\,1} + \cdots + a_{n\,n}$ of its diagonal entries, which is also equal to the sum of the eigenvalues of A. This is because the determinant of a matrix is equal to the product of its eigenvalues, and if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, by Lemma 6.4, $e^{\lambda_1}, \ldots, e^{\lambda_n}$ are the eigenvalues of e^A, and thus,

$$\det(e^A) = e^{\lambda_1} \cdots e^{\lambda_n} = e^{\lambda_1 + \cdots + \lambda_n} = e^{tr(A)}.$$

This shows that e^A is always an invertible matrix, since e^z is never null for every $z \in \mathbb{C}$. In fact, the inverse of e^A is e^{-A}, but we need to prove another Lemma. This is because it is generally not true that

$$e^{A+B} = e^A e^B,$$

unless A and B commute, i.e. $AB = BA$. We need to prove this last fact.

Lemma 6.5 *Given any two complex $n \times n$ matrices A, B, if $AB = BA$ then*

$$e^{A+B} = e^A e^B.$$

Proof. Since $AB = BA$, we can expand $(A + B)^p$ using the binomial formula:

$$(A + B)^p = \sum_{k=0}^{p} \binom{p}{k} A^k B^{p-k},$$

and thus,

$$\frac{1}{p!}(A+B)^p = \sum_{k=0}^{p} \frac{A^k B^{p-k}}{k!(p-k)!}.$$

Note that for any integer $N \geq 0$, we can write

$$\sum_{p=0}^{2N} \frac{1}{p!}(A+B)^p = \sum_{p=0}^{2N}\sum_{k=0}^{p} \frac{A^k B^{p-k}}{k!(p-k)!} = \left(\sum_{p=0}^{N} \frac{A^p}{p!}\right)\left(\sum_{p=0}^{N} \frac{B^p}{p!}\right) + \sum_{\substack{\max(k,l) > N \\ k+l \leq 2N}} \frac{A^k}{k!}\frac{B^l}{l!},$$

where there are $N(N+1)$ pairs (k, l) in the second term. Letting

$$\|A\| = \max\{|a_{i\,j}| \mid 1 \leq i, j \leq n\}, \quad \|B\| = \max\{|b_{i\,j}| \mid 1 \leq i, j \leq n\},$$

and $\mu = \max(\|A\|, \|B\|)$, note that for every entry $c_{i\,j}$ in $\frac{A^k}{k!}\frac{B^l}{l!}$, we have

$$|c_{i\,j}| \leq n\frac{(n\mu)^k}{k!}\frac{(n\mu)^l}{l!} \leq \frac{(n^2\mu)^{2N}}{N!}.$$

As a consequence, the absolute value of every entry in

$$\sum_{\substack{\max(k,l) > N \\ k+l \leq 2N}} \frac{A^k}{k!}\frac{B^l}{l!}$$

is bounded by

$$N(N+1)\frac{(n^2\mu)^{2N}}{N!},$$

which goes to 0 when $N \mapsto \infty$. From this, it immediately follows that

$$e^{A+B} = e^A e^B.$$

□

Now, using Lemma 6.5, since A and $-A$ commute, we have

$$e^A e^{-A} = e^{A+-A} = e^{0_n} = I_n,$$

which shows that the inverse of e^A is e^{-A}.

We will now use the properties of the exponential that we have just established to show how various matrices can be represented as exponentials of other matrices.

7 The Lie Groups $\mathbf{GL}(n, \mathbb{R})$, $\mathbf{SL}(n, \mathbb{R})$, $\mathbf{O}(n)$ and $\mathbf{SO}(n)$, the Lie Algebras $\mathfrak{gl}(n, \mathbb{R})$, $\mathfrak{sl}(n, \mathbb{R})$, $\mathfrak{o}(n)$, and $\mathfrak{so}(n)$, and the Exponential Map

First, we recall some basic facts and definitions. The set of real invertible $n \times n$ matrices forms a group under multiplication denoted as $\mathbf{GL}(n, \mathbb{R})$. The subset of $\mathbf{GL}(n, \mathbb{R})$ consisting of those matrices having determinant $+1$ is a subgroup of $\mathbf{GL}(n, \mathbb{R})$ denoted as $\mathbf{SL}(n, \mathbb{R})$. It is also easy to check that the set of real $n \times n$ orthogonal matrices forms a group under multiplication denoted as $\mathbf{O}(n)$. The subset of $\mathbf{O}(n)$ consisting of those matrices having determinant $+1$ is a subgroup of $\mathbf{O}(n)$ denoted as $\mathbf{SO}(n)$. We will also call matrices in $\mathbf{SO}(n)$ *rotation matrices*. Staying with easy things, we can check that the set of real $n \times n$ matrices with null trace forms a vector space under addition, and similarly for the set of skew symmetric matrices.

Definition 7.1 The group $\mathbf{GL}(n, \mathbb{R})$ is called the *general linear group*, and its subgroup $\mathbf{SL}(n, \mathbb{R})$ is called the *special linear group*. The group $\mathbf{O}(n)$ of orthogonal matrices is called the *orthogonal group*, and its subgroup $\mathbf{SO}(n)$ is called the *special orthogonal group* (or *group of rotations*). The vector space of real $n \times n$ matrices with null trace is denoted as $\mathfrak{sl}(n, \mathbb{R})$, and the vector space of real $n \times n$ skew symmetric matrices is denoted as $\mathfrak{so}(n)$.

Remark: The notations $\mathfrak{sl}(n, \mathbb{R})$ and $\mathfrak{so}(n)$ are rather strange and deserve some explanation. The groups $\mathbf{GL}(n, \mathbb{R})$, $\mathbf{SL}(n, \mathbb{R})$, $\mathbf{O}(n)$ and $\mathbf{SO}(n)$ are more than just groups. They are also topological groups, which means that they are topological spaces (viewed as subspaces of $\mathbb{R}^{n^2}$) and that the multiplication and the inverse operations are continuous (in fact, smooth), and smooth real manifolds.[2] Such objects are called *Lie groups*. The real vector spaces $\mathfrak{sl}(n)$ and $\mathfrak{so}(n)$ are what's called *Lie algebras*. However, we haven't defined the algebra structure on $\mathfrak{sl}(n, \mathbb{R})$ and $\mathfrak{so}(n)$ yet. The algebra structure is given by what's called the *Lie bracket*, which is defined as

$$[A,\, B] = AB - BA.$$

Lie algebras are associated with Lie groups. What's going on is that the Lie algebra of a Lie group is its tangent space at the identity, i.e. the space of all tangent vectors at the identity (in this case, I_n). In some sense, the Lie algebra achieves a "linearization" of the Lie group. The exponential map is a map from the Lie algebra to the Lie group, for example,

$$\exp\colon \mathfrak{so}(n) \to \mathbf{SO}(n)$$

and

$$\exp\colon \mathfrak{sl}(n, \mathbb{R}) \to \mathbf{SL}(n, \mathbb{R}).$$

[2] We refrain from defining manifolds right now, not to interupt the flow of intuitive ideas.

The exponential map often allows a parameterization of the Lie group elements by simpler objects, the Lie algebra elements.

One might ask, what happened to the Lie algebras $\mathfrak{gl}(n, \mathbb{R})$ and $\mathfrak{o}(n)$ associated with the Lie groups $\mathbf{GL}(n, \mathbb{R})$ and $\mathbf{O}(n)$? We will see later that $\mathfrak{gl}(n, \mathbb{R})$ is the set of *all* real $n \times n$ matrices, and that $\mathfrak{o}(n) = \mathfrak{so}(n)$.

The properties of the exponential map play an important role in studying a Lie group. For example, it is clear that the map

$$\exp\colon \mathfrak{gl}(n, \mathbb{R}) \to \mathbf{GL}(n, \mathbb{R})$$

is well defined, but since every matrix of the form e^A has a positive determinant, exp is not surjective. Similarly, since

$$\det(e^A) = e^{tr(A)},$$

the map

$$\exp\colon \mathfrak{sl}(n, \mathbb{R}) \to \mathbf{SL}(n, \mathbb{R})$$

is well defined. However, we showed in section 6 that it is not surjective either. As we will see in the next theorem, the map

$$\exp\colon \mathfrak{so}(n) \to \mathbf{SO}(n)$$

is well defined and surjective. The map

$$\exp\colon \mathfrak{o}(n) \to \mathbf{O}(n)$$

is well defined, but it is not surjective since there are matrices in $\mathbf{O}(n)$ with determinant -1.

Remark: The situation for matrices over the field $\mathbb{C}$ of complex numbers is quite different, as we will see later.

We now show the fundamental relationship between $\mathbf{SO}(n)$ and $\mathfrak{so}(n)$.

Theorem 7.2 *The exponential map*

$$\exp\colon \mathfrak{so}(n) \to \mathbf{SO}(n)$$

is well defined and surjective.

Proof. First, we need to prove that if A is a skew symmetric matrix, then e^A is a rotation matrix. For this, first check that

$$(e^A)^\top = e^{A^\top}.$$

Then, since $A^\top = -A$, we get

$$(e^A)^\top = e^{A^\top} = e^{-A},$$

and so

$$(e^A)^\top e^A = e^{-A}e^A = e^{-A+A} = e^{0_n} = I_n,$$

and similarly, $e^A(e^A)^\top = I_n$, showing that e^A is orthogonal. Also,

$$\det(e^A) = e^{tr(A)},$$

and since A is real skew symmetric, its diagonal entries are 0, i.e. $tr(A) = 0$, and so $\det(e^A) = +1$.

For the surjectivity, we will use Theorem 4.4 and Theorem 4.5. Theorem 4.4 says that for every skew symmetric matrix A, there is an orthogonal matrix P such that $A = PD\,P^\top$, where D is a block diagonal matrix of the form

$$D = \begin{pmatrix} D_1 & & \dots & \\ & D_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & D_p \end{pmatrix}$$

such that each block D_i is either 0 or a two-dimensional matrix of the form

$$D_i = \begin{pmatrix} 0 & -\theta_i \\ \theta_i & 0 \end{pmatrix}$$

where $\theta_i \in \mathbb{R}$, with $\theta_i > 0$. Theorem 4.5 says that for every orthogonal matrix R there is an orthogonal matrix P such that $R = PE\,P^\top$, where E is a block diagonal matrix of the form

$$E = \begin{pmatrix} E_1 & & \dots & \\ & E_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & E_p \end{pmatrix}$$

such that each block E_i is either 1, -1, or a two-dimensional matrix of the form

$$E_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

If R is a rotation matrix, there is an even number of -1, and they can be grouped into blocks of size two associated with $\theta = \pi$. Let D be the block matrix associated with E in the obvious way (where an entry 1 in E is associated with a 0 in D). Since by Lemma 6.2,

$$e^A = e^{PDP^{-1}} = Pe^DP^{-1},$$

and since D is a block diagonal matrix, we can compute e^D by computing the exponentials of its blocks. If $D_i = 0$, we get $E_i = e^0 = +1$, and if

$$D_i = \begin{pmatrix} 0 & -\theta_i \\ \theta_i & 0 \end{pmatrix}$$

we showed earlier that

$$e^{D_i} = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix},$$

exactly the block E_i. Thus, $E = e^D$, and as a consequence,

$$e^A = e^{PDP^{-1}} = Pe^D P^{-1} = PEP^{-1} = PE\,P^\top = R.$$

This shows the surjectivity of the exponential. $\square$

When $n = 3$ (and A is skew symmetric), it is possible to work out an explicit formula for e^A, known as *Rodrigues formula*. For any 3×3 real skew symmetric matrix

$$A = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix},$$

letting $\theta = \sqrt{a^2 + b^2 + c^2}$ and

$$B = \begin{pmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{pmatrix},$$

we have the following result.

Lemma 7.3 *(Rodrigues' formula (1840)). The exponential map* $\exp\colon \mathfrak{so}(3) \to \mathbf{SO}(3)$ *is given by*

$$e^A = \cos\theta\, I_3 + \frac{\sin\theta}{\theta} A + \frac{(1 - \cos\theta)}{\theta^2} B,$$

or equivalently, by

$$e^A = I_3 + \frac{\sin\theta}{\theta} A + \frac{(1 - \cos\theta)}{\theta^2} A^2$$

if $\theta \neq 0$*, with* $e^{0_3} = I_3$*.*

Proof sketch. First, prove that

$$A^2 = -\theta^2 I + B,$$
$$AB = BA = 0.$$

From the above, deduce that

$$A^3 = -\theta^2 A,$$

and for any $k \geq 0$,

$$\begin{aligned}
A^{4k+1} &= \theta^{4k} A, \\
A^{4k+2} &= \theta^{4k} A^2, \\
A^{4k+3} &= -\theta^{4k+2} A, \\
A^{4k+4} &= -\theta^{4k+2} A^2,
\end{aligned}$$

Then, prove the desired result by writing the power series for e^A, and regrouping terms so that the power series for cos and sin show up. $\square$

The above formulae are the well known formulae expressing a rotation of axis specified by the vector (a, b, c) and of angle θ. Since the exponential is surjective, it is possible to write down an explicit formula for its inverse (but it is a multi-valued function!). This has applications in kinematics, robotics, and motion interpolation.

8 Symmetric Matrices, Symmetric Positive Definite Matrices, and the Exponential Map

Recall that a real symmetric matrix is called *positive* (or *positive semi-definite*) if its eigenvalues are all positive or null, and *positive definite* if its eigenvalues are all strictly positive. We denote the vector space of real symmetric $n \times n$ matrices as $\mathbf{S}(n)$, the set of symmetric positive matrices as $\mathbf{SP}(n)$, and the set of symmetric positive definite matrices as $\mathbf{SPD}(n)$.

The next lemma shows that every symmetric positive definite matrix A is of the form e^B for some unique symmetric matrix B. The set of symmetric matrices is a vector space, but it is not a Lie algebra because the Lie bracket $[A, B]$ is not symmetric unless A and B commute, and the set of symmetric (positive) definite matrices is not a multiplicative group, so this result is of a different flavor as Theorem 7.2.

Lemma 8.1 *For every symmetric matrix B, the matrix e^B is symmetric positive definite. For every symmetric positive definite matrix A, there is a unique symmetric matrix B such that $A = e^B$.*

Proof. We showed earlier that

$$(e^B)^\top = e^{B^\top}.$$

If B is a symmetric matrix, since $B^\top = B$, we get

$$(e^B)^\top = e^{B^\top} = e^B,$$

and e^B is also symmetric. Since the eigenvalues $\lambda_1, \ldots, \lambda_n$ of the symmetric matrix B are real and the eigenvalues of e^B are $e^{\lambda_1}, \ldots, e^{\lambda_n}$, and since $e^\lambda > 0$ if $\lambda \in \mathbb{R}$, e^B is positive definite.

If A is symmetric positive definite, by Theorem 4.3, there is an orthogonal matrix P such that $A = PD\,P^\top$, where D is a diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & & \ldots & \\ & \lambda_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \lambda_n \end{pmatrix},$$

where $\lambda_i > 0$, since A is positive definite. Letting

$$L = \begin{pmatrix} \log\lambda_1 & & \ldots & \\ & \log\lambda_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \log\lambda_n \end{pmatrix},$$

it is obvious that $e^L = D$, with $\log\lambda_i \in \mathbb{R}$, since $\lambda_i > 0$. Let $B = PL\,P^\top$. By Lemma 6.2, we have

$$e^B = e^{PL\,P^\top} = e^{PLP^{-1}} = Pe^L P^{-1} = Pe^L\,P^\top = PD\,P^\top = A.$$

Finally, we prove that if B_1 and B_2 are symmetric and $A = e^{B_1} = e^{B_2}$, then $B_1 = B_2$. Since B_1 is symmetric, there is an orthonormal basis $(\overrightarrow{u_1}, \ldots, \overrightarrow{u_n})$ of eigenvectors of B_1, and let $\mu_1, \ldots, \mu_n$ be the corresponding eigenvalues. Similarly, there is an orthonormal basis $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$ of eigenvectors of B_2. We are going to prove that B_1 and B_2 agree on the basis $(\overrightarrow{v_1}, \ldots, \overrightarrow{v_n})$, thus proving that $B_1 = B_2$.

Let μ be some eigenvalue of B_2, and let $\overrightarrow{v} = \overrightarrow{v_i}$ be some eigenvector of B_2 associated with μ. We can write

$$\overrightarrow{v} = \alpha_1\overrightarrow{u_1} + \cdots + \alpha_n\overrightarrow{u_n}.$$

Since $\overrightarrow{v}$ is an eigenvector of B_2 for μ and $A = e^{B_2}$, by Lemma 6.4,

$$A(\overrightarrow{v}) = e^\mu\overrightarrow{v} = e^\mu\alpha_1\overrightarrow{u_1} + \cdots + e^\mu\alpha_n\overrightarrow{u_n}.$$

On the other hand,

$$A(\overrightarrow{v}) = A(\alpha_1\overrightarrow{u_1} + \cdots + \alpha_n\overrightarrow{u_n}) = \alpha_1 A(\overrightarrow{u_1}) + \cdots + \alpha_n A(\overrightarrow{u_n}),$$

and since $A = e^{B_1}$ and $B_1(\overrightarrow{u_i}) = \mu_i\overrightarrow{u_i}$, by Lemma 6.4, we get

$$A(\overrightarrow{v}) = e^{\mu_1}\alpha_1\overrightarrow{u_1} + \cdots + e^{\mu_n}\alpha_n\overrightarrow{u_n}.$$

Therefore, $\alpha_i = 0$ if $\mu_i \neq \mu$. Letting

$$I = \{i \mid \mu_i = \mu,\ i \in \{1, \ldots, n\}\},$$

we have

$$\overrightarrow{v} = \sum_{i \in I} \alpha_i \overrightarrow{u_i}.$$

Now,

$$B_1(\overrightarrow{v}) = B_1(\sum_{i \in I} \alpha_i \overrightarrow{u_i}) = \sum_{i \in I} \alpha_i B_1(\overrightarrow{u_i}) = \sum_{i \in I} \alpha_i \mu_i \overrightarrow{u_i} = \sum_{i \in I} \alpha_i \mu \overrightarrow{u_i} = \mu(\sum_{i \in I} \alpha_i \overrightarrow{u_i}) = \mu \overrightarrow{v},$$

since $\mu_i = \mu$ when $i \in I$. Since $\overrightarrow{v}$ is an eigenvector of B_2 for μ,

$$B_2(\overrightarrow{v}) = \mu \overrightarrow{v},$$

which shows that

$$B_1(\overrightarrow{v}) = B_2(\overrightarrow{v}).$$

Since the above holds for every eigenvector $\overrightarrow{v_i}$, we proved that $B_1 = B_2$. $\square$

Lemma 8.1 can be reformulated as stating that the map $\exp\colon \mathbf{S}(n) \to \mathbf{SPD}(n)$ is a bijection. In fact, it can be shown that it is a homeomorphism. In the case of invertible matrices, the Polar form Theorem can be reformulated as stating that there is a bijection between the topological space $\mathbf{GL}(n, \mathbb{R})$ of real $n \times n$ invertible matrices (also a group) and $\mathbf{O}(n) \times \mathbf{SPD}(n)$. As a corollary of the Polar form Theorem (Theorem 5.4) and Lemma 8.1, we have the following result: For every invertible matrix A, there is a unique orthogonal matrix R and a unique symmetric matrix S such that

$$A = R\, e^S.$$

Thus, we have a bijection between $\mathbf{GL}(n, \mathbb{R})$ and $\mathbf{O}(n) \times \mathbf{S}(n)$. But $\mathbf{S}(n)$ itself is isomorphic to $\mathbb{R}^{n(n+1)/2}$, and so, there is a bijection between $\mathbf{GL}(n, \mathbb{R})$ and $\mathbf{O}(n) \times \mathbb{R}^{n(n+1)/2}$. It can also be shown that this bijection is a homeomorphism. This is an interesting fact. Indeed, this homeomorphism essentially reduces the study of the topology of $\mathbf{GL}(n, \mathbb{R})$ to the study of the topology of $\mathbf{O}(n)$. This is nice, since it can be shown that $\mathbf{O}(n)$ is compact.

In $A = R\, e^S$, if $\det(A) > 0$, then R must a rotation matrix (i.e. $\det(R) = +1$) since $\det(e^S) > 0$. In particular, if $A \in \mathbf{SL}(n, \mathbb{R})$, since $\det(A) = \det(R) = +1$, the symmetric matrix S must have a null trace, i.e. $S \in \mathbf{S}(n) \cap \mathfrak{sl}(n, \mathbb{R})$. Thus, we have a bijection between $\mathbf{SL}(n, \mathbb{R})$ and $\mathbf{SO}(n) \times (\mathbf{S}(n) \cap \mathfrak{sl}(n, \mathbb{R}))$.

We can also use the results of section 4 to show that the exponential map is a surjective map from the skew Hermitian matrices to the unitary matrices.

9 The Lie Groups $\mathbf{GL}(n, \mathbb{C})$, $\mathbf{SL}(n, \mathbb{C})$, $\mathbf{U}(n)$ and $\mathbf{SU}(n)$, the Lie Algebras $\mathfrak{gl}(n, \mathbb{C})$, $\mathfrak{sl}(n, \mathbb{C})$, $\mathfrak{u}(n)$, and $\mathfrak{su}(n)$, and the Exponential Map

The set of complex invertible $n \times n$ matrices forms a group under multiplication denoted as $\mathbf{GL}(n, \mathbb{C})$. The subset of $\mathbf{GL}(n, \mathbb{C})$ consisting of those matrices having determinant $+1$ is a subgroup of $\mathbf{GL}(n, \mathbb{C})$ denoted as $\mathbf{SL}(n, \mathbb{C})$. It is also easy to check that the set of complex $n \times n$ unitary matrices forms a group under multiplication denoted as $\mathbf{U}(n)$. The subset of $\mathbf{U}(n)$ consisting of those matrices having determinant $+1$ is a subgroup of $\mathbf{U}(n)$ denoted as $\mathbf{SU}(n)$. We can also check that the set of complex $n \times n$ matrices with null trace forms a real vector space under addition, and similarly for the set of skew Hermitian matrices, and the set of skew Hermitian matrices with null trace.

Definition 9.1 The group $\mathbf{GL}(n, \mathbb{C})$ is called the *general linear group*, and its subgroup $\mathbf{SL}(n, \mathbb{C})$ is called the *special linear group*. The group $\mathbf{U}(n)$ of unitary matrices is called the *unitary group*, and its subgroup $\mathbf{SU}(n)$ is called the *special unitary group*. The real vector space of complex $n \times n$ matrices with null trace is denoted as $\mathfrak{sl}(n, \mathbb{C})$, the real vector space of skew Hermitian matrices is denoted as $\mathfrak{u}(n)$, and the real vector space $\mathfrak{u}(n) \cap \mathfrak{sl}(n, \mathbb{C})$ is denoted as $\mathfrak{su}(n)$.

Remark: As in the real case, the groups $\mathbf{GL}(n, \mathbb{C})$, $\mathbf{SL}(n, \mathbb{C})$, $\mathbf{U}(n)$ and $\mathbf{SU}(n)$ are also topological groups (viewed as subspaces of $\mathbb{R}^{2n^2}$), and in fact, smooth real manifolds. Such objects are called *(real) Lie groups*. The real vector spaces $\mathfrak{sl}(n, \mathbb{C})$, $\mathfrak{u}(n)$, and $\mathfrak{su}(n)$ are *Lie algebras* associated with $\mathbf{SL}(n, \mathbb{C})$, $\mathbf{U}(n)$ and $\mathbf{SU}(n)$. The algebra structure is given by the *Lie bracket*, which is defined as

$$[A, B] = AB - BA.$$

One should be very careful that even though the Lie algebras $\mathfrak{sl}(n, \mathbb{C})$, $\mathfrak{u}(n)$, and $\mathfrak{su}(n)$ consist of matrices with complex coefficients, we view them as *real* vector spaces. The Lie algebra $\mathfrak{sl}(n, \mathbb{C})$ is also a complex vector space, but $\mathfrak{u}(n)$ and $\mathfrak{su}(n)$ are not! Indeed, if A is a skew Hermitian matrix, iA is *not* skew Hermitian, but Hermitian!

Remark: It is also possible to define complex Lie groups, which means that they are topological groups and smooth *complex* manifolds. It turns out that $\mathbf{GL}(n, \mathbb{C})$ and $\mathbf{SL}(n, \mathbb{C})$ are complex manifolds, but not $\mathbf{U}(n)$ and $\mathbf{SU}(n)$.

Again the Lie algebra achieves a "linearization" of the Lie group. In the complex case, the Lie algebras $\mathfrak{gl}(n, \mathbb{C})$ is the set of *all* complex $n \times n$ matrices, but $\mathfrak{u}(n) \neq \mathfrak{su}(n)$, because a skew Hermitian matrix does not necessarily have a null trace.

The properties of the exponential map also play an important role in studying complex Lie groups. For example, it is clear that the map

$$\exp\colon \mathfrak{gl}(n,\mathbb{C}) \to \mathbf{GL}(n,\mathbb{C})$$

is well defined, but this time, it is surjective! One way to prove it is to use the Jordan normal form. Similarly, since

$$\det(e^A) = e^{tr(A)},$$

the map

$$\exp\colon \mathfrak{sl}(n,\mathbb{C}) \to \mathbf{SL}(n,\mathbb{C})$$

is well defined, but it is not surjective! As we will see in the next theorem, the maps

$$\exp\colon \mathfrak{u}(n) \to \mathbf{U}(n)$$

and

$$\exp\colon \mathfrak{su}(n) \to \mathbf{SU}(n)$$

are well defined and surjective.

Theorem 9.2 *The exponential maps*

$$\exp\colon \mathfrak{u}(n) \to \mathbf{U}(n) \quad \textit{and} \quad \exp\colon \mathfrak{su}(n) \to \mathbf{SU}(n)$$

are well defined and surjective,

Proof. First, we need to prove that if A is a skew Hermitian matrix, then e^A is a unitary matrix. For this, first check that

$$(e^A)^* = e^{A^*}.$$

Then, since $A^* = -A$, we get

$$(e^A)^* = e^{A^*} = e^{-A},$$

and so

$$(e^A)^*\, e^A = e^{-A}e^A = e^{-A+A} = e^{0_n} = I_n,$$

and similarly, $e^A\,(e^A)^* = I_n$, showing that e^A is unitary. Since

$$\det(e^A) = e^{tr(A)},$$

if A is skew Hermitian and has null trace, then $\det(e^A) = +1$.

For the surjectivity, we will use Theorem 4.7. First, assume that A is a unitary matrix. By Theorem 4.7, there is a unitary matrix U and a diagonal matrix D such that $A = UDU^*$. Furthermore, since A is unitary, the entries $\lambda_1, \ldots, \lambda_n$ in D (the eigenvalues of A) have absolute value $+1$. Thus, the entries in D are of the form $\cos\theta + i\sin\theta = e^{i\theta}$. Thus, we can assume that D is a diagonal matrix of the form

$$D = \begin{pmatrix} e^{i\theta_1} & & \dots & \\ & e^{i\theta_2} & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & e^{i\theta_p} \end{pmatrix}.$$

If we let E be the diagonal matrix

$$E = \begin{pmatrix} i\theta_1 & & \dots & \\ & i\theta_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & i\theta_p \end{pmatrix}$$

it is obvious that E is skew Hermitian and that

$$e^E = D.$$

Then, letting $B = UEU^*$, we have

$$e^B = A,$$

and it is immediately verified that B is skew Hermitian, since E is.

If A is a unitary matrix with determinant $+1$, since the eigenvalues of A are $e^{i\theta_1}, \dots, e^{i\theta_p}$ and the determinant of A is the product

$$e^{i\theta_1} \cdots e^{i\theta_p} = e^{i(\theta_1 + \cdots + \theta_p)}$$

of these eigenvalues, we must have

$$\theta_1 + \cdots + \theta_p = 0,$$

and so, E is skew Hermitian and has zero trace. As above, letting

$$B = UEU^*,$$

we have

$$e^B = A,$$

where B is skew Hermitian and has null trace. $\square$

We now extend the result of section 8 to Hermitian matrices.

10 Hermitian Matrices, Hermitian Positive Definite Matrices, and the Exponential Map

Recall that a Hermitian matrix is called *positive* (or *positive semi-definite*) if its eigenvalues are all positive or null, and *positive definite* if its eigenvalues are all strictly positive. We denote the real vector space of Hermitian $n \times n$ matrices as $\mathbf{H}(n)$, the set of Hermitian positive matrices as $\mathbf{HP}(n)$, and the set of Hermitian positive definite matrices as $\mathbf{HPD}(n)$.

The next lemma shows that every Hermitian positive definite matrix A is of the form e^B for some unique Hermitian matrix B. As in the real case, the set of Hermitian matrices is a real vector space, but it is not a Lie algebra because the Lie bracket $[A, B]$ is not Hermitian unless A and B commute, and the set of Hermitian (positive) definite matrices is not a multiplicative group.

Lemma 10.1 *For every Hermitian matrix B, the matrix e^B is Hermitian positive definite. For every Hermitian positive definite matrix A, there is a unique Hermitian matrix B such that $A = e^B$.*

Proof. It is basically the same as the proof of Theorem 10.1, except that a Hermitian matrix can be written as $A = UDU^*$, where D is a real diagonal matrix and U is unitary instead of orthogonal. $\square$

Lemma 10.1 can be reformulated as stating that the map $\exp\colon \mathbf{H}(n) \to \mathbf{HPD}(n)$ is a bijection. In fact, it can be shown that it is a homeomorphism. In the case of complex invertible matrices, the Polar form Theorem can be reformulated as stating that there is a bijection between the topological space $\mathbf{GL}(n, \mathbb{C})$ of complex $n \times n$ invertible matrices (also a group) and $\mathbf{U}(n) \times \mathbf{HPD}(n)$. As a corollary of the Polar form Theorem and Lemma 10.1, we have the following result: For every complex invertible matrix A, there is a unique unitary matrix U and a unique Hermitian matrix S such that

$$A = U\, e^S.$$

Thus, we have a bijection between $\mathbf{GL}(n, \mathbb{C})$ and $\mathbf{U}(n) \times \mathbf{H}(n)$. But $\mathbf{H}(n)$ itself is isomorphic to $\mathbb{R}^{n^2}$, and so, there is a bijection between $\mathbf{GL}(n, \mathbb{C})$ and $\mathbf{U}(n) \times \mathbb{R}^{n^2}$. It can also be shown that this bijection is a homeomorphism. This is an interesting fact. Indeed, this homeomorphism essentially reduces the study of the topology of $\mathbf{GL}(n, \mathbb{C})$ to the study of the topology of $\mathbf{U}(n)$. This is nice, since it can be shown that $\mathbf{U}(n)$ is compact (as a real manifold).

In the polar decomposition $A = Ue^S$, we have $|\det(U)| = 1$ since U is unitary, and $tr(S)$ is real since S is Hermitian (since it is the sum of the eigenvalues of S, which are real), so that $\det(e^S) > 0$. Thus, if $\det(A) = 1$, we must have $\det(e^S) = 1$, which implies that $S \in \mathbf{H}(n) \cap \mathfrak{sl}(n, \mathbb{C})$. Thus, we have a bijection between $\mathbf{SL}(n, \mathbb{C})$ and $\mathbf{SU}(n) \times (\mathbf{H}(n) \cap \mathfrak{sl}(n, \mathbb{C}))$.

In the next section, we study the group $\mathbf{SE}(n)$ of affine maps induced by orthogonal transformations, also called rigid motions, and its Lie algebra. We will show that the exponential map is surjective. The groups $\mathbf{SE}(2)$ and $\mathbf{SE}(3)$ play play a fundamental role in robotics, dynamics, and motion planning.

11 The Lie Group $\mathbf{SE}(n)$ and the Lie Algebra $\mathfrak{se}(n)$

First, we review the usual way of representing affine maps of $\mathbb{R}^n$ in terms of $(n+1)\times(n+1)$ matrices.

Definition 11.1 The set of affine maps ρ of $\mathbb{R}^n$ defined such that

$$\rho(X) = RX + U,$$

where R is a rotation matrix ($R \in \mathbf{SO}(n)$) and U is some vector in $\mathbb{R}^n$, is a group under composition called the group of *direct affine isometries, or rigid motions*, denoted as $\mathbf{SE}(n)$.

Every rigid motion can be represented by the $(n+1)\times(n+1)$ matrix

$$\begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix}$$

in the sense that

$$\begin{pmatrix} \rho(X) \\ 1 \end{pmatrix} = \begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ 1 \end{pmatrix}$$

iff

$$\rho(X) = RX + U.$$

Definition 11.2 The vector space of real $(n+1)\times(n+1)$ matrices of the form

$$A = \begin{pmatrix} \Omega & U \\ 0 & 0 \end{pmatrix},$$

where Ω is a skew symmetric matrix and U is a vector in $\mathbb{R}^n$ is denoted as $\mathfrak{se}(n)$.

Remark: The group $\mathbf{SE}(n)$ is a Lie group, and its Lie algebra turns out to be $\mathfrak{se}(n)$.

We will show that the exponential map $\exp\colon \mathfrak{se}(n) \to \mathbf{SE}(n)$ is surjective. First, we prove the following key Lemma.

Lemma 11.3 *Given any* $(n+1)\times(n+1)$ *matrix of the form*

$$A = \begin{pmatrix} \Omega & U \\ 0 & 0 \end{pmatrix}$$

where Ω *is any matrix and* $U \in \mathbb{R}^n$,

$$A^k = \begin{pmatrix} \Omega^k & \Omega^{k-1}U \\ 0 & 0 \end{pmatrix},$$

where $\Omega^0 = I_n$. *As a consequence,*

$$e^A = \begin{pmatrix} e^{\Omega} & VU \\ 0 & 1 \end{pmatrix},$$

where

$$V = I_n + \sum_{k\geq 1} \frac{\Omega^k}{(k+1)!}.$$

Proof. A trivial induction on k shows that

$$A^k = \begin{pmatrix} \Omega^k & \Omega^{k-1}U \\ 0 & 0 \end{pmatrix}.$$

Then, we have

$$\begin{aligned} e^A &= \sum_{k\geq 0} \frac{A^k}{k!}, \\ &= I_{n+1} + \sum_{k\geq 1} \frac{1}{k!} \begin{pmatrix} \Omega^k & \Omega^{k-1}U \\ 0 & 0 \end{pmatrix}, \\ &= \begin{pmatrix} I_n + \sum_{k\geq 0} \frac{\Omega^k}{k!} & \sum_{k\geq 1} \frac{\Omega^{k-1}}{k!} U \\ 0 & 1 \end{pmatrix}, \\ &= \begin{pmatrix} e^{\Omega} & VU \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

$\square$

We can now prove our main theorem. We will need to prove that V is invertible when Ω is a skew symmetric matrix. It would be tempting to write V as

$$V = \Omega^{-1}(e^{\Omega} - I).$$

Unfortunately, for odd n, a skew symmetric matrix of order n is not invertible! Thus, we have to find another way of proving that V is invertible. However, observe that we have the following useful fact:

$$V = I_n + \sum_{k\geq 1} \frac{\Omega^k}{(k+1)!} = \int_0^1 e^{\Omega t} dt.$$

This is what we will use in Theorem 11.4 to prove surjectivity.

Theorem 11.4 *The exponential map*

$$\exp\colon \mathfrak{se}(n) \to \mathbf{SE}(n)$$

is well defined and surjective.

Proof. Since Ω is skew symmetric, e^{Ω} is a rotation matrix, and by Theorem 7.2, the exponential map

$$\exp\colon \mathfrak{so}(n) \to \mathbf{SO}(n)$$

is surjective. Thus, it remains to prove that for every rotation matrix R, there is some skew symmetric matrix Ω such that $R = e^{\Omega}$ and

$$V = I_n + \sum_{k\geq 1} \frac{\Omega^k}{(k+1)!}$$

is invertible. By Theorem 4.4, for every skew symmetric matrix Ω, there is an orthogonal matrix P such that $\Omega = P D P^{\top}$, where D is a block diagonal matrix of the form

$$D = \begin{pmatrix} D_1 & & \ldots & \\ & D_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & D_p \end{pmatrix}$$

such that each block D_i is either 0 or a two-dimensional matrix of the form

$$D_i = \begin{pmatrix} 0 & -\theta_i \\ \theta_i & 0 \end{pmatrix}$$

where $\theta_i \in \mathbb{R}$, with $\theta_i > 0$. Actually, we can assume that $\theta_i \neq k2\pi$ for all $k \in \mathbb{Z}$, since when $\theta_i = k2\pi$, we have $e^{D_i} = I_2$, and D_i can be replaced by two one-dimensional blocks each consisting of a single zero. To compute V, since $\Omega = P D P^{\top} = PDP^{-1}$, observe that

$$V = I_n + \sum_{k\geq 1} \frac{\Omega^k}{(k+1)!}$$

$$
\begin{aligned}
&= I_n + \sum_{k\geq 1} \frac{PD^kP^{-1}}{(k+1)!} \\
&= P\left(I_n + \sum_{k\geq 1} \frac{D^k}{(k+1)!}\right) P^{-1} \\
&= PWP^{-1},
\end{aligned}
$$

where

$$W = I_n + \sum_{k\geq 1} \frac{D^k}{(k+1)!}.$$

We can compute

$$W = I_n + \sum_{k\geq 1} \frac{D^k}{(k+1)!} = \int_0^1 e^{Dt}dt,$$

by computing

$$W = \begin{pmatrix} W_1 & & \ldots & \\ & W_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & W_p \end{pmatrix}$$

by blocks. Since

$$e^{D_i} = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

if

$$D_i = \begin{pmatrix} 0 & -\theta_i \\ \theta_i & 0 \end{pmatrix}$$

and

$$W_i = \int_0^1 e^{D_i t}dt,$$

we get

$$W_i = \begin{pmatrix} \int_0^1 \cos(\theta_i t)dt & \int_0^1 -\sin(\theta_i t)dt \\ \int_0^1 \sin(\theta_i t)dt & \int_0^1 \cos(\theta_i t)dt \end{pmatrix} = \frac{1}{\theta_i}\begin{pmatrix} \sin(\theta_i t)\mid_0^1 & \cos(\theta_i t)\mid_0^1 \\ -\cos(\theta_i t)\mid_0^1 & \sin(\theta_i t)\mid_0^1 \end{pmatrix},$$

that is,

$$W_i = \frac{1}{\theta_i}\begin{pmatrix} \sin\theta_i & -(1-\cos\theta_i) \\ 1-\cos\theta_i & \sin\theta_i \end{pmatrix},$$

and $W_i = 1$ when $D_i = 0$. Now, in the first case where

$$W_i = \frac{1}{\theta_i}\begin{pmatrix} \sin\theta_i & -(1-\cos\theta_i) \\ 1-\cos\theta_i & \sin\theta_i \end{pmatrix},$$

the determinant is

$$\frac{1}{\theta_i^2}\left((\sin\theta_i)^2 + (1-\cos\theta_i)^2\right) = \frac{2}{\theta_i^2}(1-\cos\theta_i),$$

which is nonzero, since $\theta_i \neq k2\pi$ for all $k \in \mathbb{Z}$. Thus, each W_i is invertible, and so is W, and thus, $V = PWP^{-1}$. $\square$

In the case $n = 3$, given a skew symmetric matrix

$$\Omega = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix},$$

letting $\theta = \sqrt{a^2+b^2+c^2}$, it it easy to prove that if $\theta = 0$, then

$$e^A = \begin{pmatrix} I_3 & U \\ 0 & 1 \end{pmatrix},$$

and that if $\theta \neq 0$ (using the fact that $\Omega^3 = -\theta^2\Omega$), then

$$e^{\Omega} = I_3 + \frac{\sin\theta}{\theta}\Omega + \frac{(1-\cos\theta)}{\theta^2}\Omega^2$$

and

$$V = I_3 + \frac{(1-\cos\theta)}{\theta^2}\Omega + \frac{(\theta-\sin\theta)}{\theta^3}\Omega^2.$$

We finally reach the best vista point of our hike, the formal definition of (linear) Lie groups and Lie algebras.

12 Finale: Lie Groups and Lie Algebras

In this section, we attempt to define precisely what are Lie groups and Lie algebras. One of the reasons why Lie groups are nice is that the tangent space makes sense at any point of the group. Furthermore, the tangent space at the identity happens to have some algebra structure, that of a Lie algebra. Roughly, the tangent space at the identity provides a "linearization" of the Lie group, and it turns out that many properties of a Lie group are reflected in its Lie algebra, and that the loss of information is not too severe. The challenge that we are facing is that unless our readers are already familiar with manifolds, the amount of basic differential geometry required to define Lie groups and Lie algebras in full generality is overwhelming.

Fortunately, all the Lie groups that we need to consider are subspaces of $\mathbb{R}^N$ for some sufficiently large N. In fact, they are all isomorphic to subgroups of $\mathbf{GL}(N,\mathbb{R})$, for some suitable N, even $\mathbf{SE}(n)$, which is isomorphic to a subgroup of $\mathbf{SL}(n+1)$. Such groups are called *linear Lie groups* (or *matrix groups*). Since the groups under consideration are subspaces of $\mathbb{R}^N$, we don't need the definition of an abstract manifold. We just have to define embedded submanifolds (also called submanifolds) of $\mathbb{R}^N$ (in the case of $\mathbf{GL}(n,\mathbb{R})$, $N = n^2$). This is the path that we will follow.

In general, the difficult part in proving that a subgroup of $\mathbf{GL}(n,\mathbb{R})$ is a Lie group is to prove that it is a manifold. Fortunately, there is a charactrerization of the linear groups that obviates much of the work. This characterization rests on two theorems. First, a Lie subgroup H of a Lie group G (where H is an embedded submanifold of G) is closed in G (see Warner [26], Chapter 3, Theorem 3.21, page 97). Second, a theorem of Von Neumann and Cartan asserts that a closed subgroup of $\mathbf{GL}(n,\mathbb{R})$ is an embedded submanifold, and thus, a Lie group (see Warner [26], Chapter 3, Theorem 3.42, page 110). Thus, a linear Lie group is a closed subgroup of $\mathbf{GL}(n,\mathbb{R})$.

Since our Lie groups are subgroups (or isomorphic to subgroups) of $\mathbf{GL}(n,\mathbb{R})$ for some suitable n, it is easy to define the Lie algebra of a Lie group using curves. This approach to define the Lie algebra of a matrix group is followed by a number of authors, such as Curtis [10]. However, Curtis is rather cavalier, since he does not explain why the required curves actually exist, and thus, according to his definition, Lie algebras could be the trivial vector space! Although we will not prove the theorem of Von Neumann and Cartan, we feel that it is important to make clear why the definitions make sense, i.e., why we are not dealing with trivial objects.

A small annoying technical problem will arise in our approach, the problem with discrete subgroups. If A is a subset of $\mathbb{R}^N$, recall that A inherits a topology from $\mathbb{R}^N$ called the *subspace topology*, and defined such that a subset V of A is open if

$$V = A \cap U$$

for some open subset U of $\mathbb{R}^N$. A point $a \in A$ is said to be *isolated* if there is there is some open subset U of $\mathbb{R}^N$ such that

$$\{a\} = A \cap U,$$

in other words, if $\{a\}$ is an open set in A.

The group $\mathbf{GL}(n,\mathbb{R})$ of real invertible $n \times n$ matrices can be viewed as a subset of $\mathbb{R}^{n^2}$, and as such, it is a topological space under the subspace topology (in fact, a dense open subset of $\mathbb{R}^{n^2}$). One can easily check that multiplication and the inverse operation are continuous, and in fact smooth (i.e., C^∞-continuously differentiable). This makes $\mathbf{GL}(n,\mathbb{R})$ a *topological group*. Any subgroup G of $\mathbf{GL}(n,\mathbb{R})$ is also a topological space under the subspace topology. A subgroup G is called a *discrete subgroup* if it has some isolated point. This turns out to be equivalent to the fact that every point of G is isolated, and thus, G has the discrete topology (every subset of G is open). Now, because $\mathbf{GL}(n,\mathbb{R})$ is Hausdorff, it can be shown that every discrete subgroup of $\mathbf{GL}(n,\mathbb{R})$ is closed (which means that its complement is open). Thus, discrete subgroups of $\mathbf{GL}(n,\mathbb{R})$ are Lie groups! But these are not very interesting Lie groups, and so, we will only consider closed subgroups of $\mathbf{GL}(n,\mathbb{R})$ that are not discrete.

Let us now review the definition of an embedded submanifold. For more details, see DoCarmo [11, 12], Marsden and Ratiu [17], Berger and Gostiaux [4], or Warner [26]. To simplify the terminology, we will use the terminology manifold (but other authors would say embedded submanifolds, or something like that).

The intuition behind the notion of a manifold in $\mathbb{R}^N$ is that a subspace M is a manifold of dimension m if every point $p \in M$ is contained in some open subset set U of M (in the subspace topology) that can be parameterized by some function $\varphi\colon \Omega \to U$ from some open subset Ω of the origin in $\mathbb{R}^m$, and that φ has some nice properties that allow the definition of smooth functions on M, and the definition of the tangent space at p. For this, φ has to be at least a homeomorphism, but more is needed: φ must be smooth and the derivative $\varphi'(0_m)$ at the origin must be injective (letting $0_m = \underbrace{(0, \ldots, 0)}_{m}$).

Definition 12.1 Given any integers N, m, with $N \geq m \geq 1$, an *m-dimensional smooth manifold in $\mathbb{R}^N$, for short, a manifold*, is a nonempty subset M of $\mathbb{R}^N$ such that for every point $p \in M$, there are two open subsets $\Omega \subseteq \mathbb{R}^m$ and $U \subseteq M$, with $p \in U$, and a smooth function $\varphi\colon \Omega \to \mathbb{R}^N$ such that φ is a homeomorphism between Ω and $U = \varphi(\Omega)$, and $\varphi'(t_0)$ is injective, where $t_0 = \varphi^{-1}(p)$. The function $\varphi\colon \Omega \to U$ is called a *(local) parameterization of M at p*. If $0_m \in \Omega$ and $\varphi(0_m) = p$, we say that $\varphi\colon \Omega \to U$ is *centered at p*.

Recall that $M \subseteq \mathbb{R}^N$ is a topological space under the subspace topology, and U is some open subset of M in the subspace topology, which means that $U = M \cap W$ for some open subset W of $\mathbb{R}^N$. Since $\varphi\colon \Omega \to U$ is a homeomorphism, it has an inverse $\varphi^{-1}\colon U \to \Omega$ which is also a homeomorphism, called a *(local) chart*. Since $\Omega \subseteq \mathbb{R}^m$, for every point $p \in M$, for every parameterization $\varphi\colon \Omega \to U$ of M at p, we have $\varphi^{-1}(p) = (z_1, \ldots, z_m)$ for some $z_i \in \mathbb{R}$, and we call $z_1, \ldots, z_m$ the *local coordinates of p (w.r.t. φ^{-1})*. We often refer to a manifold M without explicitly specifying its dimension (the integer m).

Intuitively, a chart provides a "flatened" local map of a region on a manifold. For instance, in the case of surfaces (2-dimensional manifolds), a chart is analogous to a planar map of a region on the surface. For a concrete example, consider a map giving a planar representation of a country, a region on the earth, a curved surface.

Remark: We could allow $m = 0$ in definition 12.1. If so, a manifold of dimension 0 is just a set of isolated points, and thus, it has the discrete topology. In fact, it can be shown that a discrete subset of $\mathbb{R}^N$ is countable. Such manifolds are not very exciting, but they do correspond to discrete subgroups.

As an example of a 2-manifold, the unit sphere S^2 in $\mathbb{R}^3$ defined such that

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$$

is a smooth 2-manifold, because it can be parameterized using the following two maps φ_1 and φ_2:

$$\varphi_1\colon (u, v) \mapsto \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1}\right)$$

and

$$\varphi_2\colon (u,v) \mapsto \left(\frac{2u}{u^2+v^2+1}, \frac{2v}{u^2+v^2+1}, \frac{1-u^2-v^2}{u^2+v^2+1} \right).$$

The map φ_1 corresponds to the inverse of the stereographic projection from the North pole $N = (0,0,1)$ onto the plane $z = 0$, and the map φ_2 corresponds to the inverse of the stereographic projection from the South pole $S = (0,0,-1)$ onto the plane $z = 0$, as illustrated in Figure 1.

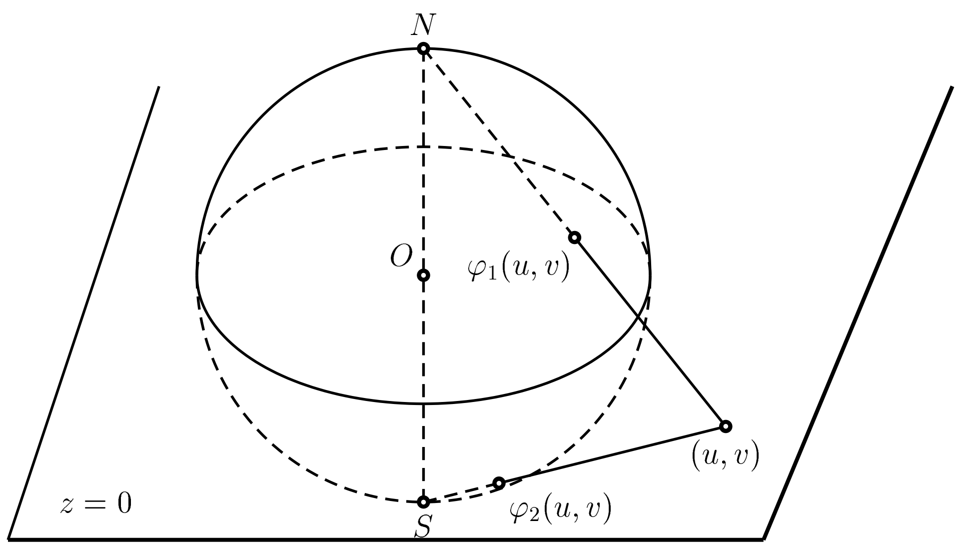

Figure 1: Inverse stereographic projections

We leave as an exercise to check that the map φ_1 parameterizes $S^2 - \{N\}$ and that the map φ_2 parameterizes $S^2 - \{S\}$ (and that they are smooth, homeomorphisms, etc). Using φ_1, the open lower hemisphere is parameterized by the open disk of center O and radius 1 contained in the plane $z = 0$. The chart φ^{-1} assigns local coordinates to the points in the open lower hemisphere. If we draw a grid of coordinate lines parallel to the x and y axes inside the open unit disk and map these lines onto the lower hemisphere using φ_1, we get curved lines on the lower hemisphere. These "coordinate lines" on the lower hemisphere provide local coordinates for every point on the lower hemisphere. For this reason, older books often talk about *curvilinear coordinate systems*, to mean the coordinate lines on a surface induced by a chart. We urge our readers to define a manifold structure on a torus. This can be done using four charts.

Every open subset of $\mathbb{R}^N$ is a manifold in a trivial way. Indeed, we can use the inclusion map as a parameterization. In particular, $\mathbf{GL}(n,\mathbb{R})$ is an open subset of $\mathbb{R}^{n^2}$, since its complement is closed (the set of invertible matrices is the inverse image of the determinant function, which is continuous). Thus, $\mathbf{GL}(n,\mathbb{R})$ is a manifold. We can view $\mathbf{GL}(n,\mathbb{C})$ as a subset of $\mathbb{R}^{(2n)^2}$ using the embedding defined as follows: for every complex $n \times n$ matrix A, construct the real $2n \times 2n$ matrix such that every entry $a + ib$ in A is replaced by the 2×2 block

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

where $a, b \in \mathbb{R}$. It is immediately verified that this map is in fact a group isomorphism. Thus, we can view $\mathbf{GL}(n, \mathbb{C})$ as a subgroup of $\mathbf{GL}(2n, \mathbb{R})$, and as a manifold in $\mathbb{R}^{(2n)^2}$.

A 1-manifold is called a *(smooth) curve*, and a 2-manifold is called a *(smooth) surface* (although some authors require that they are also connected).

The following two lemmas provide the link with the definition of an abstract manifold. The first lemma is easily shown using the inverse function theorem.

Lemma 12.2 *Given an m-dimensional manifold M in $\mathbb{R}^N$, for every $p \in M$, there are two open sets $\Omega, W \subseteq \mathbb{R}^N$ with $0_N \in \Omega$ and $p \in M \cap W$, and a smooth diffeomorphism $\varphi\colon \Omega \to W$, such that $\varphi(0_N) = p$ and*

$$\varphi(\Omega \cap (\mathbb{R}^m \times \{0_{N-m}\})) = M \cap W.$$

The next lemma is easily shown from Lemma 12.2. It is a key technical result used to show that interesting properties of maps between manifolds do not depend on parameterizations.

Lemma 12.3 *Given an m-dimensional manifold M in $\mathbb{R}^N$, for every $p \in M$, for any two parameterizations $\varphi_1\colon \Omega_1 \to U_1$ and $\varphi_2\colon \Omega_2 \to U_2$ of M at p, if $U_1 \cap U_2 \neq \emptyset$, the map $\varphi_2^{-1} \circ \varphi_1\colon \varphi_1^{-1}(U_1 \cap U_2) \to \varphi_2^{-1}(U_1 \cap U_2)$ is a smooth diffeomorphism.*

The maps $\varphi_2^{-1} \circ \varphi_1\colon \varphi_1^{-1}(U_1 \cap U_2) \to \varphi_2^{-1}(U_1 \cap U_2)$ are called *transitions maps*. Lemma 12.3 is illustrated in Figure 2.

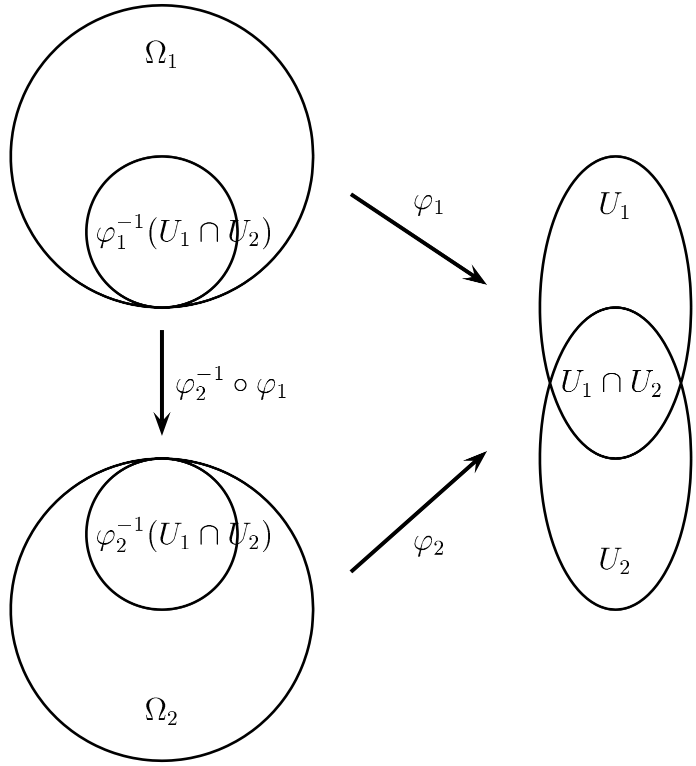

Figure 2: Parameterizations and transition functions

Let us review the definition of a smooth curve in a manifold and of the tangent vector at a point of a curve.

Definition 12.4 Let M be an m-dimensional manifold in $\mathbb{R}^N$. A *smooth curve* γ *in* M is any function $\gamma: I \to M$, where I is an open interval in $\mathbb{R}$, and such that for every $t \in I$, letting $p = \gamma(t)$, there is some parameterization $\varphi: \Omega \to U$ of M at p and some open interval $]t - \epsilon,\ t + \epsilon[\subseteq I$ such that the curve $\varphi^{-1} \circ \gamma:]t - \epsilon,\ t + \epsilon[\to \mathbb{R}^m$ is smooth.

Using Lemma 12.3, it is easily shown that Definition 12.4 does not depend on the choice of the parameterization $\varphi: \Omega \to U$ at p.

Lemma 12.3 also implies that γ viewed as a curve $\gamma: I \to \mathbb{R}^N$ is smooth. Then, the *tangent vector to the curve* $\gamma: I \to \mathbb{R}^N$ *at* t, denoted as $\gamma'(t)$, is the value of the derivative of γ at t (a vector in $\mathbb{R}^N$) computed as usual:

$$\gamma'(t) = \lim_{h \mapsto 0} \frac{\gamma(t + h) - \gamma(t)}{h}.$$

Given any point $p \in M$, we will show that the set of tangent vectors to all smooth curves in M through p is a vector space isomorphic to the vector space $\mathbb{R}^m$. The tangent vector at p to a curve γ on a manifold M is illustrated in Figure 3.

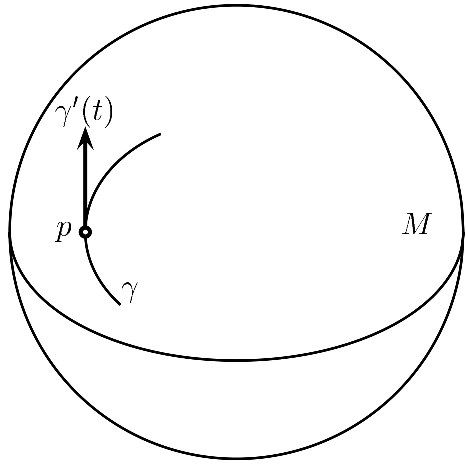

Figure 3: Tangent vector to a curve on a manifold

Given a smooth curve $\gamma: I \to M$, for any $t \in I$, letting $p = \gamma(t)$, since M is a manifold, there is a parameterization $\varphi: \Omega \to U$ such that $\varphi(0_m) = p \in U$ and some open interval $J \subseteq I$ with $t \in J$, such that the function

$$\varphi^{-1} \circ \gamma: J \to \mathbb{R}^m$$

is a smooth curve, since γ is a smooth curve. Letting $\alpha = \varphi^{-1} \circ \gamma$, the derivative $\alpha'(t)$ is well defined, and it is a vector in $\mathbb{R}^m$. But $\varphi \circ \alpha: J \to M$ is also a smooth curve which agrees with γ on J, and by the chain rule,

$$\gamma'(t) = \varphi'(0_m)(\alpha'(t)),$$

since $\alpha(t) = 0_m$ (because $\varphi(0_m) = p$ and $\gamma(0) = p$). Observe that $\gamma'(t)$ is a vector in $\mathbb{R}^N$. Now, for every vector $\overrightarrow{v} \in \mathbb{R}^m$, the curve $\alpha: I \to \mathbb{R}^m$ defined such that

$$\alpha(u) = (u - t)\overrightarrow{v}$$

for all $u \in I$ is clearly smooth, and $\alpha'(t) = \overrightarrow{v}$. This shows that the set of tangent vectors at t to all smooth curves (in $\mathbb{R}^m$) passing through 0_m is the entire vector space $\mathbb{R}^m$. Since every smooth curve $\gamma: I \to M$ agrees with a curve of the form $\varphi \circ \alpha: J \to M$ for some smooth curve $\alpha: J \to \mathbb{R}^m$ (with $J \subseteq I$) as explained above, and since it is assumed that $\varphi'(0_m)$ is injective, $\varphi'(0_m)$ maps the vector space $\mathbb{R}^m$ injectively to the set of tangent vectors to γ at p, as claimed. All this is summarized in the following definition.

Definition 12.5 Let M be an m-dimensional manifold in $\mathbb{R}^N$. For every point $p \in M$, the *tangent space* T_pM *at* p is the set of all vectors in $\mathbb{R}^N$ of the form $\gamma'(0)$, where $\gamma: I \to M$ is any smooth curve in M such that $p = \gamma(0)$. The set T_pM is a vector space isomorphic to $\mathbb{R}^m$. Every vector $\overrightarrow{v} \in T_pM$ is called a *tangent vector to* M *at* p.

We can now define Lie groups.

Definition 12.6 A *Lie group* is a nonempty subset G of $\mathbb{R}^N$ $(N \geq 1)$ satisfying the following conditions:

(a) G is a group.

(b) G is a manifold in $\mathbb{R}^N$.

(c) The group operation $\cdot : G \times G \to G$ and the inverse map ${}^{-1}\colon G \to G$ are smooth.

It is immediately verified that $\mathbf{GL}(n, \mathbb{R})$ is a Lie group. Since all the Lie groups that we are considering are subgroups of $\mathbf{GL}(n, \mathbb{R})$, the following definition is in order.

Definition 12.7 A *linear Lie group* is a subgroup G of $\mathbf{GL}(n, \mathbb{R})$ (for some $n \geq 1$) which is also a smooth manifold in $\mathbb{R}^{n^2}$.

Let $\mathbf{M}(n, \mathbb{R})$ denote the set of all real $n \times n$ matrices (invertible or not). If we recall that the exponential map

$$\exp\colon A \mapsto e^A$$

is well defined on $\mathbf{M}(n, \mathbb{R})$, we have the following crucial theorem due to Von Neumann and Cartan.

Theorem 12.8 *A closed subgroup G of $\mathbf{GL}(n, \mathbb{R})$ is a linear Lie group. Furthermore, the set $\mathfrak{g}$ defined such that*

$$\mathfrak{g} = \{X \in \mathbf{M}(n, \mathbb{R}) \mid e^{tX} \in G \text{ for all } t \in \mathbb{R}\}$$

is a vector space equal to the tangent space $T_I G$ at the identity I, and $\mathfrak{g}$ is closed under the Lie bracket $[-, -]$ defined such that $[A, B] = AB - BA$ for all $A, B \in \mathbf{M}(n, \mathbb{R})$.

Theorem 12.8 applies even when G is a discrete subgroup, but in this case, $\mathfrak{g}$ is trivial (i.e., $\mathfrak{g} = \{0\}$). For example, the set of nonnull reals $\mathbb{R}^* = \mathbb{R} - \{0\} = \mathbf{GL}(1, \mathbb{R})$ is a Lie group under multiplication, and the subgroup

$$H = \{2^n \mid n \in \mathbb{Z}\}$$

is a discrete subgroup of $\mathbb{R}^*$. Thus, H is a Lie group. On the other hand, the set $\mathbb{Q}^* = \mathbb{Q} - \{0\}$ of nonnull rational numbers is a multiplicative subgroup of $\mathbb{R}^*$, but is it not closed since $\mathbb{Q}$ is dense in $\mathbb{R}$.

The proof of theorem 12.8 involves proving that when G is not a discrete subgroup, there is an open subset $\Omega \subseteq \mathbf{M}(n, \mathbb{R})$ such that $0_{n,n} \in \Omega$, an open subset of $W \subseteq \mathbf{M}(n, \mathbb{R})$ such that $I \in W$, and that $\exp\colon \Omega \to W$ is a diffeomorphism such that

$$\exp(\Omega \cap \mathfrak{g}) = W \cap G.$$

If G is closed and not discrete, we must have $m \geq 1$, and $\mathfrak{g}$ has dimension m.

With the help of Theorem 12.8, it is now very easy to prove that $\mathbf{SL}(n)$, $\mathbf{O}(n)$, $\mathbf{SO}(n)$, $\mathbf{SL}(n, \mathbb{C})$, $\mathbf{U}(n)$, and $\mathbf{SU}(n)$, are Lie groups. We can also prove that $\mathbf{SE}(n)$ is a Lie group as follows. Recall that we can view every element of $\mathbf{SE}(n)$ as a real $(n+1) \times (n+1)$ matrix

$$\begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix}$$

where $R \in \mathbf{SO}(n)$ and $U \in \mathbb{R}^n$. In fact, such matrices belong to $\mathbf{SL}(n+1)$. This embedding of $\mathbf{SE}(n)$ into $\mathbf{SL}(n+1)$ is a group homomorphism, since the the group operation on $\mathbf{SE}(n)$ corresponds to

$$\begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix} \begin{pmatrix} S & V \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} RS & RV + U \\ 0 & 1 \end{pmatrix}.$$

and the inverse to

$$\begin{pmatrix} R^{-1} & -R^{-1}U \\ 0 & 1 \end{pmatrix}.$$

Also note that the embedding shows that, as a manifold, $\mathbf{SE}(n)$ is diffeomorphic to $\mathbf{SO}(n) \times R^n$ (given a manifold M_1 of dimension m_1 and a manifold M_2 of dimension m_2, the product $M_1 \times M_2$ can be given the structure of a manifold of dimension $m_1 + m_2$ in a natural way). Thus, $\mathbf{SE}(n)$ is a Lie group whith underlying manifold $\mathbf{SO}(n) \times R^n$, and in fact, a subgroup of $\mathbf{SL}(n+1)$.

Even though $\mathbf{SE}(n)$ is diffeomorphic to $\mathbf{SO}(n) \times R^n$ as a manifold, it is *not* isomorphic to $\mathbf{SO}(n) \times R^n$ as a group, because the group multiplication on $\mathbf{SE}(n)$ is not the multiplication on $\mathbf{SO}(n) \times R^n$. Instead, $\mathbf{SE}(n)$ is a *semidirect product* of $\mathbf{SO}(n)$ and R^n.

Going back to Theorem 12.8, the vector space $\mathfrak{g}$ is called the *Lie algebra* of the Lie group G. Lie algebras are defined as follows.

Definition 12.9 A *(real) Lie algebra* $\mathcal{A}$ is a real vector space together with a bilinear map $[\cdot, \cdot] \colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ called the *Lie bracket* on $\mathcal{A}$, such that the following two identities hold for all $a, b, c \in \mathcal{A}$:

$$[a,\, a] = 0,$$

and the so-called *Jacobi identity*:

$$[a,\, [b,\, c]] + [c,\, [a,\, b]] + [b,\, [c,\, a]] = 0.$$

It is immediately verified that $[b, a] = -[a, b]$. In view of Theorem 12.8, the vector space $\mathfrak{g} = T_I G$ associated with a Lie group G is indeed a Lie algebra. Furthermore, the exponential map $\exp\colon \mathfrak{g} \to G$ is well defined. In general, exp is neither injective nor surjective, as we observed earlier. Theorem 12.8 also provides a kind of recipe for "computing" the Lie algebra $\mathfrak{g} = T_I G$ of a Lie group G. Indeed, $\mathfrak{g}$ is the tangent space to G at I, and thus, we can use curves to compute tangent vectors. Actually, for every $X \in T_I G$, the map

$$\gamma_X\colon t \mapsto e^{tX}$$

is a smooth curve in G, and it is easily shown that $\gamma_X'(0) = X$. Thus, we can use these curves. As an illustration, we show that the Lie algebras of $\mathbf{SL}(n)$ and $\mathbf{SO}(n)$ are the matrices with null trace and the skew symmetric matrices.

Let $t \mapsto R(t)$ be a smooth curve in $\mathbf{SL}(n)$ such that $R(0) = I$. We have $\det(R(t)) = 1$ for all $t \in]-\epsilon, \epsilon[$. Using the chain rule, we can compute the derivative of the function

$$t \mapsto \det(R(t))$$

at $t = 0$, and we get

$$\det{}'_I(R'(0)) = 0.$$

It is an easy exercise to prove that

$$\det{}'_I(X) = tr(X),$$

and thus, $tr(R'(0)) = 0$, which says that the tangent vector $X = R'(0)$ has null trace. Another proof consists in observing that $X \in \mathfrak{sl}(n, \mathbb{R})$ iff

$$\det(e^{tX}) = 1$$

for all $t \in \mathbb{R}$. Since $\det(e^{tX}) = e^{tr(tX)}$, for $t = 1$, we get $tr(X) = 0$, as claimed. Clearly, $\mathfrak{sl}(n, \mathbb{R})$ has dimension $n^2 - 1$.

Let $t \mapsto R(t)$ be a smooth curve in $\mathbf{SO}(n)$ such that $R(0) = I$. Since each $R(t)$ is orthogonal, we have

$$R(t)\, R(t)^\top = I$$

for all $t \in]-\epsilon, \epsilon[$. Taking the derivative at $t = 0$, we get

$$R'(0)\, R(0)^\top + R(0)\, R'(0)^\top = 0,$$

but since $R(0) = I = R(0)^\top$, we get

$$R'(0) + R'(0)^\top = 0,$$

which says that the tangent vector $X = R'(0)$ is skew symmetric. Since the diagonal elements of a skew symmetric matrix are null, the trace is automatically null, and the condition

$\det(R) = 1$ yields nothing new. This shows that $\mathfrak{o}(n) = \mathfrak{so}(n)$. It is easily shown that $\mathfrak{so}(n)$ has dimension $n(n-1)/2$.

As a concrete example, the Lie algebra $\mathfrak{so}(3)$ of $\mathbf{SO}(3)$ is the real vector space consisting of all 3×3 real skew symmetric matrices. Every such matrix is of the form

$$\begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix}$$

where $b, c, d \in \mathbb{R}$. The Lie bracket $[A, B]$ in $\mathfrak{so}(3)$ is also given by the usual commutator, $[A, B] = AB - BA$.

We can define an isomorphism of Lie algebras $\psi\colon (\mathbb{R}^3, \times) \to \mathfrak{so}(3)$ by the formula

$$\psi(b, c, d) = \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix}.$$

It is indeed easy to verify that

$$\psi(\overrightarrow{u} \times \overrightarrow{v}) = [\psi(\overrightarrow{u}), \psi(\overrightarrow{v})].$$

It is also easily verified that for any two vectors $\overrightarrow{u} = (b, c, d)$ and $\overrightarrow{v} = (b', c', d')$ in $\mathbb{R}^3$,

$$\psi(\overrightarrow{u})(\overrightarrow{v}) = \overrightarrow{u} \times \overrightarrow{v}.$$

The exponential map $\exp\colon \mathfrak{so}(3) \to \mathbf{SO}(3)$ is given by Rodrigues' formula (see Lemma 7.3):

$$e^A = \cos\theta\, I_3 + \frac{\sin\theta}{\theta} A + \frac{(1-\cos\theta)}{\theta^2} B,$$

or equivalently by

$$e^A = I_3 + \frac{\sin\theta}{\theta} A + \frac{(1-\cos\theta)}{\theta^2} A^2$$

if $\theta \neq 0$, where

$$A = \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix},$$

$\theta = \sqrt{b^2 + c^2 + d^2}$, $B = A^2 + \theta^2 I_3$, and with $e^{0_3} = I_3$.

Using the above methods, it is easy to verify that the Lie algebras $\mathfrak{gl}(n, \mathbb{R})$, $\mathfrak{sl}(n, \mathbb{R})$, $\mathfrak{o}(n)$ and $\mathfrak{so}(n)$, are respectively $\mathbf{M}(n, \mathbb{R})$, the set of matrices with null trace, and the set of skew symmetric matrices (in the last two cases). A similar computation can be done for $\mathfrak{gl}(n, \mathbb{C})$, $\mathfrak{sl}(n, \mathbb{C})$, $\mathfrak{u}(n)$ and $\mathfrak{su}(n)$, confirming the claims of Section 9. It is easy to show that $\mathfrak{gl}(n, \mathbb{C})$

has dimension $2n^2$, $\mathfrak{sl}(n,\mathbb{C})$ has dimension $2(n^2-1)$, $\mathfrak{u}(n)$ has dimension n^2, and that $\mathfrak{su}(n)$ has dimension n^2-1.

For example, the Lie algebra $\mathfrak{su}(2)$ of $\mathbf{SU}(2)$ (or S^3) is the real vector space consisting of all 2×2 (complex) skew Hermitian matrices of null trace. Every such matrix is of the form

$$i(d\sigma_1 + c\sigma_2 + b\sigma_3) = \begin{pmatrix} ib & c+id \\ -c+id & -ib \end{pmatrix},$$

where $b, c, d \in \mathbb{R}$, and $\sigma_1, \sigma_2, \sigma_3$ are the Pauli spin matrices, and thus, the matrices $i\sigma_1, i\sigma_2, i\sigma_3$ form a basis of the Lie algebra $\mathfrak{su}(2)$. The Lie bracket $[A, B]$ in $\mathfrak{su}(2)$ is given by the usual commutator, $[A, B] = AB - BA$.

It is easily checked that the vector space $\mathbb{R}^3$ is a Lie algebra if we define the Lie bracket on $\mathbb{R}^3$ as the usual cross-product $\overrightarrow{u} \times \overrightarrow{v}$ of vectors. Then, we can define an isomorphism of Lie algebras $\varphi\colon (\mathbb{R}^3, \times) \to \mathfrak{su}(2)$ by the formula

$$\varphi(b, c, d) = \frac{i}{2}(d\sigma_1 + c\sigma_2 + b\sigma_3) = \frac{1}{2}\begin{pmatrix} ib & c+id \\ -c+id & -ib \end{pmatrix}.$$

It is indeed easy to verify that

$$\varphi(\overrightarrow{u} \times \overrightarrow{v}) = [\varphi(\overrightarrow{u}),\, \varphi(\overrightarrow{v})].$$

Going back to $\mathfrak{su}(2)$, letting $\theta = \sqrt{b^2 + c^2 + d^2}$, we can write

$$d\sigma_1 + c\sigma_2 + b\sigma_3 = \begin{pmatrix} b & -ic+d \\ ic+d & -b \end{pmatrix} = \theta A,$$

where

$$A = \frac{1}{\theta}(d\sigma_1 + c\sigma_2 + b\sigma_3) = \frac{1}{\theta}\begin{pmatrix} b & -ic+d \\ ic+d & -b \end{pmatrix},$$

so that $A^2 = I$, and it can be shown that the exponential map $\exp\colon \mathfrak{su}(2) \to \mathbf{SU}(2)$ is given by

$$\exp(i\theta A) = \cos\theta\, \mathbf{1} + i \sin\theta\, A.$$

In view of the isomorphism $\varphi\colon (\mathbb{R}^3, \times) \to \mathfrak{su}(2)$, where

$$\varphi(b, c, d) = \frac{1}{2}\begin{pmatrix} ib & c+id \\ -c+id & -ib \end{pmatrix} = i\frac{\theta}{2}A,$$

the exponential map can be viewed as a map $\exp\colon (\mathbb{R}^3, \times) \to \mathbf{SU}(2)$ given by the formula

$$\exp(\theta v) = \left[\cos\frac{\theta}{2},\, \sin\frac{\theta}{2}\, v\right],$$

for every vector θv, where v is a unit vector in $\mathbb{R}^3$, and $\theta \in \mathbb{R}$. In this form, $\exp(\theta v)$ is a quaternion corresponding to a rotation of axis v and angle θ.

As we showed, $\mathbf{SE}(n)$ is a Lie group, and its lie algebra $\mathfrak{se}(n)$ described in section 11 is easily determined as the subalgebra of $\mathfrak{sl}(n+1)$ consisting of all matrices of the form

$$\begin{pmatrix} B & U \\ 0 & 0 \end{pmatrix}$$

where $B \in \mathfrak{so}(n)$ and $U \in \mathbb{R}^n$. Thus, $\mathfrak{se}(n)$ has dimension $n(n+1)/2$. The Lie bracket is given by

$$\begin{pmatrix} B & U \\ 0 & 0 \end{pmatrix}\begin{pmatrix} C & V \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} C & V \\ 0 & 0 \end{pmatrix}\begin{pmatrix} B & U \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} BC - CB & BV - CU \\ 0 & 0 \end{pmatrix}.$$

We conclude by indicating the relationship between homomorphisms of Lie groups and homomorphisms of Lie algebras. First, we need to explain what is a smooth map between manifolds.

Definition 12.10 Let M_1 and M_2 be m_1-dimensional and m_2-dimensional manifolds in $\mathbb{R}^N$. A function $f\colon M_1 \to M_2$ is *smooth* if for every $p \in M_1$, there are parameterizations $\varphi\colon \Omega_1 \to U_1$ of M_1 at p and $\psi\colon \Omega_2 \to U_2$ of M_2 at $f(p)$, such that $f(U_1) \subseteq U_2$ and

$$\psi^{-1} \circ f \circ \varphi\colon \Omega_1 \to \mathbb{R}^{m_2}$$

is smooth.

Using Lemma 12.3, it is easily shown that Definition 12.10 does not depend on the choice of the parameterizations $\varphi\colon \Omega_1 \to U_1$ and $\psi\colon \Omega_2 \to U_2$. A smooth map f between manifolds is a *smooth diffeomorphism* if f is bijective and both f and f^{-1} are smooth maps.

We now define the derivative of a smooth map between manifolds.

Definition 12.11 Let M_1 and M_2 be m_1-dimensional and m_2-dimensional manifolds in $\mathbb{R}^N$. For any smooth function $f\colon M_1 \to M_2$, for any $p \in M_1$, the function $f'_p\colon T_pM_1 \to T_{f(p)}M_2$ called the *tangent map of f at p, or derivative of f at p, or differential of f at p*, is defined as follows: for every $\overrightarrow{v} \in T_pM_1$, for every smooth curve $\gamma\colon I \to M_1$ such that $\gamma(0) = p$ and $\gamma'(0) = \overrightarrow{v}$,

$$f'_p(\overrightarrow{v}) = (f \circ \gamma)'(0).$$

The map f'_p is also denoted as df_p, or T_pf. Doing a few calculations involving the facts that

$$f \circ \gamma = (f \circ \varphi) \circ (\varphi^{-1} \circ \gamma) \quad \text{and} \quad \gamma = \varphi \circ (\varphi^{-1} \circ \gamma)$$

and using Lemma 12.3, it is not hard to show that $f'_p(\overrightarrow{v})$ does not depend on the choice of the curve γ. It is easily shown that f'_p is a linear map.

Finally, we define homomorphisms of Lie groups and Lie algebras and see how they relate.

Definition 12.12 Given two Lie groups G_1 and G_2, a *homomorphism (or map) of Lie groups* is a function $f\colon G_1 \to G_2$ which is a homomorphism of groups and a smooth map (between the manifolds G_1 and G_2). Given two Lie algebras $\mathcal{A}_1$ and $\mathcal{A}_2$, a *homomorphism (or map) of Lie algebras* is a function $f\colon \mathcal{A}_1 \to \mathcal{A}_2$ which is a linear map between the vector spaces $\mathcal{A}_1$ and $\mathcal{A}_2$ and which preserves Lie brackets, i.e.,

$$f([A, B]) = [f(A), f(B)]$$

for all $A, B \in \mathcal{A}_1$.

An *isomorphism of Lie groups* is a bijective function f such that both f and f^{-1} are maps of Lie groups, and an *isomorphism of Lie algebras* is a bijective function f such that both f and f^{-1} are maps of Lie algebras. It is immediately verified that if $f\colon G_1 \to G_2$ is a homomorphism of Lie groups, then $f'_I\colon \mathfrak{g}_1 \to \mathfrak{g}_2$ is a homomorphism of Lie algebras. If some additional assumptions are made about G_1 and G_2, (for example, connected, simply-connected), it can be shown that f is pretty much determined by f'_I.

Alert readers must have noticed that we only defined the Lie algebra of a linear group. In the more general case, we can still define the Lie algebra $\mathfrak{g}$ of a Lie group G as the tangent space $T_I G$ at the identity I. The tangent space $\mathfrak{g} = T_I G$ is a vector space, but we need to define the Lie bracket. This can be done in several ways. We explain briefly how this can be done in terms of so-called adjoint representations. This has the advantage of not requiring the definition of left-invariant vector fields, but it is still a little bizarre!

Given a Lie group G, for every $a \in G$, we define *left translation* as the map $L_a\colon G \to G$ such that $L_a(b) = ab$ for all $b \in G$, and *right translation* as the map $R_a\colon G \to G$ such that $R_a(b) = ba$ for all $b \in G$. The maps L_a and R_a are diffeomorphisms, and their derivatives play an important role. The inner automorphisms $R_{a^{-1}} \circ L_a$ (also written as $R_{a^{-1}} L_a$) also plays an important role. Note that

$$R_{a^{-1}} L_a(b) = aba^{-1}.$$

The derivative

$$(R_{a^{-1}} L_a)'_I\colon \mathfrak{g} \to \mathfrak{g}$$

of $R_{a^{-1}} L_a$ at I is an isomorphism of Lie algebras denoted as $Ad_a\colon \mathfrak{g} \to \mathfrak{g}$. The map $a \mapsto Ad_a$ is a map of Lie groups

$$Ad\colon G \to \mathbf{GL}(\mathfrak{g}),$$

called the *adjoint representation of* G (where $\mathbf{GL}(\mathfrak{g})$ denotes the Lie group of all bijective linear maps on $\mathfrak{g}$).

In the case of a linear group, one can verify that

$$Ad(a)(X) = Ad_a(X) = aXa^{-1}$$

for all $a \in G$ and all $X \in \mathfrak{g}$. The derivative

$$Ad'_I \colon \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$$

of Ad at I is map of Lie algebras denoted as $ad \colon \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$, called the *adjoint representation of* $\mathfrak{g}$ (where $\mathfrak{gl}(\mathfrak{g})$ denotes the Lie algebra of all linear maps on $\mathfrak{g}$).

In the case of a linear group, it can be verified that

$$ad(A)(B) = [A,\, B]$$

for all $A, B \in \mathfrak{g}$. One can also check that the Jacobi identity on $\mathfrak{g}$ is equivalent to the fact that ad preserves Lie brackets, i.e., ad is a map of Lie algebras:

$$ad([A,\, B]) = [ad(A),\, ad(B)]$$

for all $A, B \in \mathfrak{g}$ (where on the right, the Lie bracket is the commutator of linear maps on $\mathfrak{g}$). Thus, we recover the Lie bracket from ad.

This is the key to the definition of the Lie bracket in the case of a general Lie group (not just a linear Lie group). We define the Lie bracket on $\mathfrak{g}$ as

$$[A,\, B] = ad(A)(B).$$

To be complete, we would have to define the exponential map $\exp \colon \mathfrak{g} \to G$ for a general Lie group. For this, we would need to introduce some left-invariant vector fields induced by the derivatives of the left translations, and integral curves associated with such vector fields.

This is not hard, but we feel that it is now time to stop our introduction to Lie groups and Lie algebras, even though we have not even touched many important topics, for instance, vector fields and differential foms. Readers who wish to learn more about Lie groups and Lie algebras should consult (more or less listed in order of difficulty) Curtis [10], Sattinger and Weaver [20], and Marsden and Ratiu [17]. Classics such as Weyl [27] and Chevalley [9] are definitely worth consulting, although the presentation and the terminology may seem a bit old fashion. Some applications of Lie groups and Lie algebra are discussed in Selig [21], and especially Murray, Li, and Sastry [19]. For more advanced texts, one may consult Abraham and Marsden [1], Warner [26], Sternberg [22], and Bröcker and tom Dieck [7]. For those who read French, Mneimné and Testard [18] is very clear and quite thorough, and uses very little differential geometry, although it is more advanced that Curtis. Chapter 1 by Bryant in Freed and Uhlenbeck [8] is also worth reading, but the pace is fast, and Chapters 7 and 8 of Fulton and Harris [13] are very good, but familiarity with manifolds is assumed.

References

[1] Ralph Abraham and Jerrold E. Marsden. *Foundations of Mechanics*. Addison Wesley, second edition, 1978.

[2] Michael Artin. *Algebra*. Prentice Hall, first edition, 1991.

[3] Marcel Berger. *Géométrie 1*. Nathan, 1990. English edition: Geometry 1, Universitext, Springer Verlag.

[4] Marcel Berger and Bernard Gostiaux. *Géométrie différentielle: variétés, courbes et surfaces*. Collection Mathématiques. Puf, second edition, 1992. English edition: Differential geometry, manifolds, curves, and surfaces, GTM No. 115, Springer Verlag.

[5] J.E. Bertin. *Algèbre linéaire et géométrie classique*. Masson, first edition, 1981.

[6] Nicolas Bourbaki. *Algèbre, Chapitre 9*. Eléments de Mathématiques. Hermann, 1968.

[7] T. Bröcker and T. tom Dieck. *Representation of Compact Lie Groups*. GTM, Vol. 98. Springer Verlag, first edition, 1985.

[8] R.L. Bryant. An introduction to Lie groups and symplectic geometry. In D.S. Freed and K.K. Uhlenbeck, editors, *Geometry and Quantum Field Theory*, pages 5–181. AMS, Providence, Rhode Island, 1995.

[9] Claude Chevalley. *Theory of Lie Groups I*. Princeton mathmetical Series, No. 8. Princeton University Press, eight printing edition, 1946.

[10] Morton L. Curtis. *Matrix Groups*. Universitext. Springer Verlag, second edition, 1984.

[11] Manfredo P. do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice Hall, 1976.

[12] Manfredo P. do Carmo. *Riemannian Geometry*. Birkauser, second edition, 1992.

[13] William Fulton and Joe Harris. *Representation Theory, A first course*. GTM No. 129. Springer Verlag, first edition, 1991.

[14] Roger Howe. Very basic Lie theory. *American Mathematical Monthly*, 90:600–623, 1983.

[15] Serge Lang. *Algebra*. Addison Wesley, third edition, 1993.

[16] Saunders Mac Lane and Garrett Birkhoff. *Algebra*. Macmillan, first edition, 1967.

[17] J.E. Marsden and T.S. Ratiu. *Introduction to Mechanics and Symmetry*. TAM, Vol. 17. Springer Verlag, first edition, 1994.

[18] R. Mneimné and F. Testard. *Introduction à la Théorie des Groupes de Lie Classiques*. Hermann, first edition, 1997.

[19] R.M. Murray, Z.X. Li, and S.S. Sastry. *A Mathematical Introduction to Robotics Manipulation*. CRC Press, first edition, 1994.

[20] D.H. Sattinger and O.L. Weaver. *Lie Groups and Algebras with Applications to Physics, Geometry, and Mechanics*. Applied Math. Science, Vol. 61. Springer Verlag, first edition, 1986.

[21] J.M. Selig. *Geometrical Methods In Robotics*. Monographs In Computer Science. Springer Verlag, first edition, 1996.

[22] S. Sternberg. *Lectures On Differential Geometry*. AMS Chelsea, second edition, 1983.

[23] G.W. Stewart. On the early history of the singular value decomposition. *SIAM review*, 35(4):551–566, 1993.

[24] Gilbert Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, first edition, 1986.

[25] Gilbert Strang. *Linear Algebra and its Applications*. Saunders HBJ, third edition, 1988.

[26] Frank Warner. *Foundations of Differentiable Manifolds and Lie Groups*. GTM No. 94. Springer Verlag, first edition, 1983.

[27] Hermann Weyl. *The Classical Groups. Their Invariants and Representations*. Princeton Mathematical Series, No. 1. Princeton University Press, second edition, 1946.