Multilinear Algebra and Its Applications

Ana Cannas, Özlem Imamoglu, Alessandra Iozzi

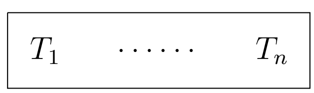
March 2021

Contents

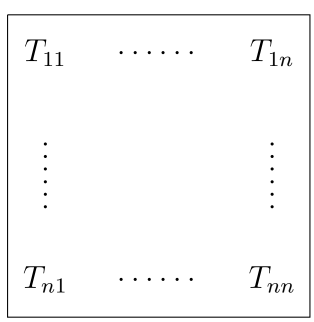
Introduction

This text deals with physical or geometric entities, known as **tensors**, which can be thought of as a generalization of vectors. The quantitative description of tensors, i.e., their description in terms of numbers, changes when we change the *frame of reference*, a.k.a. the *basis* in linear algebra. Tensors are central in Engineering and Physics, because they provide the framework for formulating and solving problems in areas such as Mechanics (inertia tensor, stress tensor, elasticity tensor, etc.), Electrodynamics (electrical conductivity and electrical resistivity tensors, electromagnetic tensor, magnetic susceptibility tensor, etc.), or General Relativity (stress–energy tensor, curvature tensor, etc.).

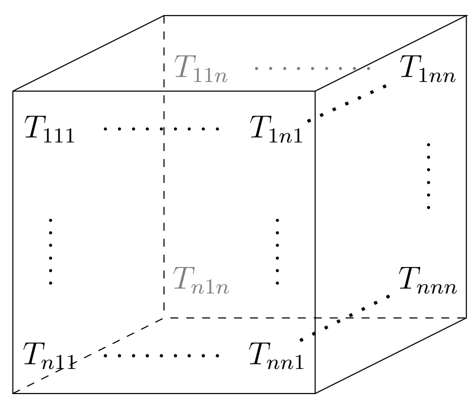
Just like the main protagonists in Linear Algebra are *vectors* and *linear maps*, the main protagonists in Multilinear Algebra are *tensors* and *multilinear maps*. Tensors describe linear relations among objects in space, and are represented – once a basis is chosen – by *multidimensional arrays of numbers*:



1. A tensor of order 1, T_j.



2. A tensor of order 2, T_{ij}.



3. A tensor of order 3, T_{ijk}.

In the notation, the indices can be upper or lower. For tensors of order at least 2, some indices can be upper and some lower. The numbers in the arrays are called **components** of the tensor and give the representation of the tensor *with respect to a given basis*.

Two natural questions arise:

(1) Why do we need tensors?
(2) What are the important features of tensors?

(1) Scalars are not enough to describe directions, for which we need to resort to vectors. At the same time, vectors might not be enough, in that they lack the ability to "modify" vectors.

EXAMPLE 0.1. We denote by **B** the magnetic fluid density measured in $\mathrm{V}\cdot\mathrm{s}/\mathrm{m}^2$ and by $\mathbf{H}$ the magnetizing intensity measured in A/m.[1] They are related by the formula

$$\mathbf{B} = \mu \mathbf{H}\,,$$

where μ is the permeability of the medium in H/m. In free space, $\mu = \mu_0 = 4\pi \times 10^{-7}\,\mathrm{H}/\mathrm{m}$ is a scalar, so that the flux density and the magnetization are vectors that differ only by their magnitude.

Other material however have properties that make these terms differ both in magnitude and direction. In such materials the scalar permeability is replaced by the tensor permeability $\boldsymbol{\mu}$ and

$$\mathbf{B} = \boldsymbol{\mu} \cdot \mathbf{H}\,.$$

Being vectors, **B** and **H** are tensors of order 1, and $\boldsymbol{\mu}$ is a tensor of order 2. We will see that they are of different type, and in fact *the order of* **H** *"cancels out" with the order of* $\boldsymbol{\mu}$ *to give a tensor of order 1.* □

(2) Physical laws do not change with different coordinate systems, hence tensors describing them must satisfy some *invariance* properties. While tensors remain intrinsically invariant with respect to changes of bases, their components will vary according to two fundamental modes: **covariance** and **contravariance**, depending on whether the components change in a way parallel to the change of basis or in an opposite way.

Here is an example of a familiar tensor from Linear Algebra, illustrating the effect of the change of basis.

EXAMPLE 0.2. We recall here the transformation property that vectors enjoy according to which they are an example of a **contravariant tensor of first order**. We use here freely notions and properties that will be recalled in the next chapter.

Let $\mathcal{B} = \{b_1, b_2, b_3\}$ and $\widetilde{\mathcal{B}} = \{\tilde{b}_1, \tilde{b}_2, \tilde{b}_3\}$ be two basis of a vector space V. A vector $v \in V$ can be written as

$$v = v^1 b_1 + v^2 b_2 + v^3 b_3\,,$$

or

$$v = \tilde{v}^1 \tilde{b}_1 + \tilde{v}^2 \tilde{b}_2 + \tilde{v}^3 \tilde{b}_3\,,$$

where v^1, v^2, v^3 (resp. $\tilde{v}^1, \tilde{v}^2, \tilde{v}^3$) are the coordinate of v with respect to the basis $\mathcal{B}$ (resp. $\widetilde{\mathcal{B}}$).

[1] The physical units here are: Volt V, second s, meter m, Ampere A, Henry H.

Warning: *Please keep the lower indices as lower indices and the upper ones as upper ones. You will see later that there is a reason for it!*

We use the following notation:

$$[v]_{\mathcal{B}} = \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} \quad \text{and} \quad [v]_{\widetilde{\mathcal{B}}} = \begin{pmatrix} \tilde{v}^1 \\ \tilde{v}^2 \\ \tilde{v}^3 \end{pmatrix}, \tag{0.1}$$

and we are interested in finding the relation between the coordinates of v in the two bases.

The vectors $\tilde{b}_j$, $j = 1, 2, 3$, in the basis $\widetilde{\mathcal{B}}$ can be written as a linear combination of vectors in $\mathcal{B}$ as follows:

$$\tilde{b}_j = L^1_j b_1 + L^2_j b_2 + L^3_j b_3 \,,$$

for some $L^i_j \in \mathbb{R}$. We consider the matrix of the change of basis from $\mathcal{B}$ to $\widetilde{\mathcal{B}}$,

$$L := L_{\widetilde{\mathcal{B}}\mathcal{B}} = \begin{bmatrix} L^1_1 & L^1_2 & L^1_3 \\ L^2_1 & L^2_2 & L^2_3 \\ L^3_1 & L^3_2 & L^3_3 \end{bmatrix}$$

whose jth-column consists of the coordinates of the vectors $\tilde{b}_j$ with respect to the basis $\mathcal{B}$. The equalities

$$\begin{cases} \tilde{b}_1 = L^1_1 b_1 + L^2_1 b_2 + L^3_1 b_3 \\ \tilde{b}_2 = L^1_2 b_1 + L^2_2 b_2 + L^3_2 b_3 \\ \tilde{b}_3 = L^1_3 b_1 + L^2_3 b_2 + L^3_3 b_3 \end{cases}$$

can simply be written as

$$\begin{pmatrix} \tilde{b}_1 & \tilde{b}_2 & \tilde{b}_3 \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} L \,. \tag{0.2}$$

(Check this symbolic equation using the rules of matrix multiplication.) Analogously, writing basis vectors in a row and vector coordinates in a column, we can write

$$v = v^1 b_1 + v^2 b_2 + v^3 b_3 = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} \tag{0.3}$$

and

$$v = \tilde{v}^1 \tilde{b}_1 + \tilde{v}^2 \tilde{b}_2 + \tilde{v}^3 \tilde{b}_3 = \begin{pmatrix} \tilde{b}_1 & \tilde{b}_2 & \tilde{b}_3 \end{pmatrix} \begin{pmatrix} \tilde{v}^1 \\ \tilde{v}^2 \\ \tilde{v}^3 \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} L \begin{pmatrix} \tilde{v}^1 \\ \tilde{v}^2 \\ \tilde{v}^3 \end{pmatrix}, \tag{0.4}$$

where we used (0.2) in the last equality. Comparing the expression of v in (0.3) and in (0.4), we conclude that

$$L\begin{pmatrix}\tilde{v}^1\\ \tilde{v}^2\\ \tilde{v}^3\end{pmatrix} = \begin{pmatrix}v^1\\ v^2\\ v^3\end{pmatrix}$$

or equivalently

$$\begin{pmatrix}\tilde{v}^1\\ \tilde{v}^2\\ \tilde{v}^3\end{pmatrix} = L^{-1}\begin{pmatrix}v^1\\ v^2\\ v^3\end{pmatrix}$$

We say that the components of a vector v are *contravariant* because they change by L^{-1} when the basis changes by L; see Section 1.3.2. A vector v is hence a *contravariant 1-tensor* or *tensor of order* $(1,0)$. □

Example 0.3 (A numerical example). Let

$$\mathcal{E} = \{e_1, e_2, e_3\} = \left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\} \tag{0.5}$$

be the *standard* basis or $\mathbb{R}^3$ and let

$$\widetilde{\mathcal{B}} = \{\tilde{b}_1, \tilde{b}_2, \tilde{b}_3\} = \left\{\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}, \begin{bmatrix}7\\8\\0\end{bmatrix}\right\}$$

be another basis of $\mathbb{R}^3$. The vector[2] $v = \begin{bmatrix}1\\1\\1\end{bmatrix}$ has coordinates

$$[v]_{\mathcal{B}} = \begin{pmatrix}1\\1\\1\end{pmatrix} \qquad \text{and} \qquad [v]_{\widetilde{\mathcal{E}}} = \begin{pmatrix}-\frac{1}{3}\\ \frac{1}{3}\\ 0\end{pmatrix}.$$

[2] For a general basis $\mathcal{B}$, the notation $[\,\cdot\,]_{\mathcal{B}}$ indicates the "operation" of taking the vector v and looking at its coordinates in the basis $\mathcal{B}$. However, in order to "write down explicitly" a vector (that is three real numbers that we write in column), one needs to give coordinates and the coordinates are usually given with respect to the standard basis. In this case there is the slightly confusing fact that

$$\begin{bmatrix}v^1\\ \vdots\\ v^n\end{bmatrix} = v = [v]_{\mathcal{E}} = \begin{pmatrix}v^1\\ \vdots\\ v^n\end{pmatrix}.$$

Since it is easy to check that

$$\begin{cases} \tilde{b}_1 = 1 \cdot e_1 + 2 \cdot e_2 + 3 \cdot e_3 \\ \tilde{b}_2 = 4 \cdot e_1 + 5 \cdot e_2 + 6 \cdot e_3 \\ \tilde{b}_3 = 7 \cdot e_1 + 8 \cdot e_2 + 0 \cdot e_3 \end{cases},$$

the matrix of the change of coordinates from $\mathcal{E}$ to $\widetilde{\mathcal{B}}$ is

$$L = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 0 \end{bmatrix} .$$

It is easy to check that

$$\begin{pmatrix} -\frac{1}{3} \\ \frac{1}{3} \\ 0 \end{pmatrix} = L^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

or equivalently

$$L \begin{pmatrix} -\frac{1}{3} \\ \frac{1}{3} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} .$$

□

CHAPTER 1

Review of Linear Algebra

1.1. Vector Spaces

A *vector space* (or *linear space*) is a set of objects where addition and *scaling* are defined in a way that satisfies natural requirements for such operations, such as the properties listed in the definition below.

1.1.1. Vectors and Scalars.

In this text, we will only consider *real* vector spaces, a.k.a. vector spaces over $\mathbb{R}$, where the scaling is by real numbers.

DEFINITION 1.1. A **vector space** V over $\mathbb{R}$ is a set V equipped with two operations:

(1) *Vector addition:* $V \times V \to V$, $(v, w) \mapsto v + w$, and
(2) *Multiplication by a scalar:* $\mathbb{R} \times V \to V$, $(\alpha, v) \mapsto \alpha v$,

satisfying the following properties:

(1) (associativity) $(u + v) + w = u + (v + w)$ for every $u, v, w \in V$;
(2) (commutativity) $u + v = v + u$ for every $u, v \in V$;
(3) (existence of the zero vector) There exists $0 \in V$ such that $v + 0 = v$ for every $v \in V$;
(4) (existence of additive inverse) For every $v \in V$, there exists $w_v \in V$ such that $v + w_v = 0$. The vector w_v is denoted by $-v$.
(5) $\alpha(\beta v) = (\alpha\beta)v$ for every $\alpha, \beta \in \mathbb{R}$ and every $v \in \mathbb{R}$;
(6) $1v = v$ for every $v \in V$;
(7) $\alpha(u + w) = \alpha u + \alpha v$ for all $\alpha \in \mathbb{R}$ and $u, v \in V$;
(8) $(\alpha + \beta)v = \alpha v + \beta v$ for all $\alpha, \beta \in \mathbb{R}$ and $v \in V$.

An element of the vector space is called a **vector** and, mostly in the context of vector spaces, a real number is called a **scalar**.

EXAMPLE 1.2 (Prototypical example). The Euclidean space $\mathbb{R}^n$, $n = 1, 2, 3, \ldots$, is a vector space with componentwise addition and multiplication by scalars. Vectors in $\mathbb{R}^n$ are denoted by

$$v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

with $x_1, \dots, x_n \in \mathbb{R}$. Addition component-by-component translates geometrically to the *parallelogram law* for vector addition, well-known in $\mathbb{R}^2$ and $\mathbb{R}^3$. □

EXAMPLES 1.3 (Other examples). The operations of vector addition and scalar multiplication are usually inferred from the context.

(1) The set of real polynomials of degree $\leq n$ is a vector space, denoted by
$$V = \mathbb{R}[x]_n := \{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 : a_j \in \mathbb{R}\}$$
with the usual (degreewise) sum of polynomials and scalar multiplication.

(2) The set of real matrices of size $m \times n$,
$$V = M_{m\times n}(\mathbb{R}) := \left\{ \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix} : a_{ij} \in \mathbb{R} \right\}$$
with componentwise addition and scalar multiplication.

(3) The space $\{f : W \to \mathbb{R}\}$ of all real-valued functions on a vector space W.

(4) The space of solutions of a homogeneous linear (ordinary or partial) differential equation.

□

EXERCISE 1.4. Are the following vector spaces?

(1) The set V of all vectors in $\mathbb{R}^3$ perpendicular to the vector $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

(2) The set of invertible 2×2 matrices, that is
$$V := \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : ad - bc = \neq 0 \right\}.$$

(3) The set of polynomials of degree exactly n, that is
$$V := \{a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n : a_j \in \mathbb{R}, a_n \neq 0\}.$$

(4) The set V of 2×4 matrices with last column zero, that is
$$V := \left\{ \begin{bmatrix} a & b & c & 0 \\ d & e & f & 0 \end{bmatrix} : a, b, c, d, e, f \in \mathbb{R} \right\}$$

(5) The set of solutions $f : \mathbb{R} \to \mathbb{R}$ of the equation $f' = 5$, that is
$$V := \{f : \mathbb{R} \to \mathbb{R} : f(x) = 5x + C, C \in \mathbb{R}\}.$$

DEFINITION 1.5. A function $T : V \to W$ between real vector spaces V and W is a **linear transformation** if it satisfies the property
$$T(\alpha v + \beta w) = \alpha T(v) + \beta T(w),$$
for all $\alpha, \beta \in \mathbb{R}$ and all $v, w \in V$.

EXERCISE 1.6. Show that the set of all linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^3$ forms a vector space.

1.1.2. Subspaces.

DEFINITION 1.7. A subset W of a vector space V that is itself a vector space is a **subspace**.

By reviewing the properties in the definition of vector space, we see that a subset $W \subseteq V$ is a subspace exactly when the following conditions are verified:

(1) The 0 element is in W;
(2) W is *closed under addition*, that is $v + w \in W$ for every $v, w \in W$;
(3) W is *closed under multiplication by scalars*, that is $\alpha v \in W$ for every $\alpha \in \mathbb{R}$ and every $v \in W$.

Condition (1) in fact follows from (2) and (3) under the assumption that $W \neq \emptyset$. Yet it is often an easy way to check that a subset is not a subspace.

Recall that a **linear combination** of vectors $v_1, \ldots, v_n \in V$ is a vector of the form $\alpha_1 v_1 + \cdots + \alpha_n v_n$ for $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. With this notion, the above three conditions for a subspace are equivalent to the following ones:

(1)' W is nonempty;
(2)' W is *closed under linear combinations*, that is $\alpha v + \beta w \in W$ for all $\alpha, \beta \in \mathbb{R}$ and all $v, w \in W$.

DEFINITION 1.8. If $T : V \to W$ is a linear transformation between real vector spaces V and W, then:

- the **kernel** (or *null space*) of T is the set $\ker T := \{v \in V : T(v) = 0\}$;
- the **image** (or *range*) of T is the set $\operatorname{im} T := \{T(v) : v \in V\}$.

EXERCISE 1.9. Show that, for a linear transformation $T : V \to W$, the kernel $\ker T$ is a subspace of V and the image $\operatorname{im} T$ is a subspace of W.

1.2. Bases

The key to study and to compute in vector spaces is the concept of *basis*, which in turn relies on the fundamental notions of *linear independence/dependence* and of *span*.

1.2.1. Definition of Basis.

DEFINITION 1.10. The vectors $b_1, \ldots, b_n \in V$ are **linearly independent** if $\alpha_1 b_1 + \cdots + \alpha_n b_n = 0$ implies that $\alpha_1 = \cdots = \alpha_n = 0$. In other words, if the only linear combination of these vectors that yields the zero vector is the trivial one. We then also say that the vector set $\{b_1, \ldots, b_n\}$ is *linearly independent*.

EXAMPLE 1.11. The vectors

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

are linearly independent in $\mathbb{R}^3$. In fact,

$$\mu_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + \mu_2 \begin{bmatrix}0\\1\\0\end{bmatrix} + \mu_3 \begin{bmatrix}0\\0\\1\end{bmatrix} = 0 \Longleftrightarrow \begin{bmatrix}\mu_1\\\mu_2\\\mu_3\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix} \Longleftrightarrow \mu_1 = \mu_2 = \mu_3 = 0\,.$$

□

EXAMPLE 1.12. The vectors

$$b_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}, \qquad b_2 = \begin{bmatrix}4\\5\\6\end{bmatrix}, \qquad b_3 = \begin{bmatrix}7\\8\\0\end{bmatrix}$$

are linearly independent in $\mathbb{R}^3$. In fact,

$$\mu_1 b_1 + \mu_2 b_2 + \mu_3 b_3 = 0 \Longleftrightarrow \begin{cases}\mu_1 + 4\mu_2 + 7\mu_3 = 0\\ 2\mu_1 + 5\mu_2 + 8\mu_3 0\\ 3\mu_1 + 6\mu_2 = 0\end{cases} \Longleftrightarrow \cdots \Longleftrightarrow \mu_1 = \mu_2 = \mu_3 = 0\,.$$

(If you are unsure how to fill in the dots look at Example 1.21.) □

EXAMPLE 1.13. The vectors

$$b_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}, \qquad b_2 = \begin{bmatrix}4\\5\\6\end{bmatrix}, \qquad b_3 = \begin{bmatrix}7\\8\\9\end{bmatrix}$$

are linearly *dependent* in $\mathbb{R}^3$, i.e., not linearly independent. In fact,

$$\mu_1 b_1 + \mu_2 b_2 + \mu_3 b_3 = 0 \Longleftrightarrow \begin{cases}\mu_1 + 4\mu_2 + 7\mu_3 = 0\\ 2\mu_1 + 5\mu_2 + 8\mu_3 0\\ 3\mu_1 + 6\mu_2 + 9\mu_3 = 0\end{cases} \Longleftrightarrow \cdots \Longleftrightarrow \begin{cases}\mu_1 = \mu_2\\ \mu_2 = -2\mu_3\,,\end{cases}$$

so

$$b_1 - 2b_2 + b_3 = 0$$

and b_1, b_2, b_3 are not linearly independent. For instance, we say that $b_1 = 2b_2 - b_3$ is a *non-trivial linear relation* between the vectors b_1, b_2 and b_3. □

DEFINITION 1.14. The vectors $b_1, \dots, b_n \in V$ **span** V, if every vector $v \in V$ can be written as a linear combination $v = \alpha_1 b_1 + \cdots + \alpha_n b_n$, for some $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. We then also say that the vector set $\{b_1, \dots, b_n\}$ *spans* V.

EXAMPLES 1.15.

(1) The vectors $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$ span $\mathbb{R}^3$.

(2) The vectors $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}$ also span $\mathbb{R}^3$.

(3) The vectors $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}$ span the xy-coordinate plane (i.e., the subspace given by the equation $z = 0$) in $\mathbb{R}^3$.

□

EXERCISE 1.16. The set of all linear combinations of $b_1, \dots, b_n \in V$ is denoted by $\operatorname{span}\{b_1, \dots, b_n\}$. Show that $\operatorname{span}\{b_1, \dots, b_n\}$ is a subspace of V.

DEFINITION 1.17. The vectors $b_1, \dots, b_n \in V$ form a **basis** of V, if:

(1) they are *linearly independent* and
(2) they *span* V.

We then denote this basis as an *ordered* set $\mathcal{B} := \{b_1, \dots, b_n\}$, where we fix the order of the vectors.

Warning: *In this text, we only consider bases for so-called* **finite-dimensional** *vector spaces, that is, vector spaces that admit bases consisting of a finite number of elements.*

EXAMPLE 1.18. The vectors

$$e_1 := \begin{bmatrix}1\\0\\0\end{bmatrix}, \qquad e_2 := \begin{bmatrix}0\\1\\0\end{bmatrix}, \qquad e_3 := \begin{bmatrix}0\\0\\1\end{bmatrix}$$

form a basis of $\mathbb{R}^3$. This is called the **standard basis** of $\mathbb{R}^3$ and denoted $\mathcal{E} := \{e_1, e_2, e_3\}$. For $\mathbb{R}^n$, the *standard basis* $\mathcal{E} := \{e_1, \dots, e_n\}$ is defined similarly. □

EXAMPLE 1.19. The vectors in Example 1.12 span $\mathbb{R}^3$, while the vectors in Example 1.13 do not span $\mathbb{R}^3$. To see this, we recall the following facts about bases. □

1.2.2. Facts about Bases.

Let V be a vector space; as stated above, V is *finite-dimensional*. Then we have:

(1) All bases of V have the same number of elements. This number is called the **dimension** of V and denoted $\dim V$.
(2) If $\mathcal{B} = \{b_1, \dots, b_n\}$ is a basis of V, there is a *unique* way of writing any $v \in V$ as a linear combination

$$v = v^1 b_1 + \dots v^n b_n$$

of elements in $\mathcal{B}$. The numbers $v^1, \dots, v^n$ are the **coordinates** of v with respect to the basis $\mathcal{B}$ and we denote by

$$[v]_{\mathcal{B}} = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$$

the **coordinate vector** of v with respect to $\mathcal{B}$.

(3) If we know that $\dim V = n$, then:
   (a) More than n vectors in V must be linearly dependent;
   (b) Fewer than n vectors in V cannot span V;
   (c) Any n linearly independent vectors span V;
   (d) Any n vectors that span V must be linearly independent;
   (e) If k vectors span V, then $k \geq n$ and some subset of those k vectors must be a basis of V;
   (f) If a set of m vectors is linearly independent, then $m \leq n$ and we can always complete that set to form a basis of V.

EXAMPLE 1.20. The vectors b_1, b_2, b_3 in Example 1.12 form a basis of $\mathbb{R}^3$ since they are linearly independent and they are exactly as many as the dimension of $\mathbb{R}^3$. □

EXAMPLE 1.21 (Gauss-Jordan elimination). We are going to compute here the coordinates of $v = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ with respect to the basis $\mathcal{B} := \{b_1, b_2, b_3\}$ from Example 1.12. The seeked coordinates $[v]_{\mathcal{B}} = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$ must satisfy the equation

$$v^1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + v^2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + v^3 \begin{bmatrix} 7 \\ 8 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

so to find them we have to solve the following system of linear equations:

$$\begin{cases} v^1 + 4v^2 + 7v^3 = 1 \\ 2v^1 + 5v^2 + 8v^3 = 1 \\ 3v^1 + 6v^2 \qquad\quad = 1. \end{cases}$$

For that purpose, we may equivalently reduce the following augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 4 & 7 & 1 \\ 2 & 5 & 8 & 1 \\ 3 & 6 & 0 & 1 \end{array}\right]$$

to echelon form, using the Gauss–Jordan elimination method. We are going to perform both calculations in parallel, which will also point out that they are indeed seemingly different incarnations of the same method.

By multiplying the first equation/row by 2 (resp. 3) and subtracting it from the second (reps. third) equation/row we obtain

$$\begin{cases} v^1 + 4v^2 + 7v^3 = 1 \\ \quad - 3v^2 - 6v^3 = -1 \\ \quad - 6v^2 - 21v^3 = -2 \end{cases} \qquad \leftrightsquigarrow \qquad \left[\begin{array}{ccc|c} 1 & 4 & 7 & 1 \\ 0 & -3 & -6 & -1 \\ 0 & -6 & -21 & -2 \end{array}\right].$$

By multiplying the second equation/row by $-\frac{1}{3}$ and by adding to the first (resp. third) equation/row the second equation/row multiplied by $-\frac{4}{3}$ (resp. 2) we obtain

$$\begin{cases} v^1 \qquad - v^3 = -\frac{1}{3} \\ \qquad v^2 + 2v^3 = \frac{1}{3} \\ \qquad\quad - 9v^3 = 0 \end{cases} \qquad \rightsquigarrow \qquad \left[\begin{array}{ccc|c} 1 & 0 & -1 & -\frac{1}{3} \\ 0 & 1 & 2 & \frac{1}{3} \\ 0 & 0 & -9 & 0 \end{array}\right].$$

The last equation/row shows that $v^3 = 0$, hence by backward substitution we obtain the solution

$$\begin{cases} v^1 \qquad\qquad = -\frac{1}{3} \\ \qquad v^2 \qquad = \frac{1}{3} \\ \qquad\qquad v^3 = 0 \end{cases} \qquad \rightsquigarrow \qquad \left[\begin{array}{ccc|c} 1 & 0 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{array}\right].$$

□

Example 1.22. When $V = \mathbb{R}^n$ and $\mathcal{B} = \mathcal{E}$ is the standard basis, the coordinate vector $v \in \mathbb{R}^n$ coincides with the vector itself! In this very special case, we have $[v]_{\mathcal{E}} = v$. □

Exercise 1.23. Let V be the vector space consisting of all 2×2 matrices with trace zero, namely

$$V := \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a, b, c, d \in \mathbb{R} \text{ and } a + d = 0 \right\}.$$

(1) Show that

$$\mathcal{B} := \left\{ \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}}_{b_1}, \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{b_2}, \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}}_{b_3} \right\}$$

is a basis of V.

(2) Show that

$$\widetilde{\mathcal{B}} := \left\{ \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}}_{\tilde{b}_1}, \underbrace{\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}}_{\tilde{b}_2}, \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{\tilde{b}_3} \right\}$$

is another basis of V.

(3) Compute the coordinates of

$$v = \begin{bmatrix} 2 & 1 \\ 7 & -2 \end{bmatrix}$$

with respect to $\mathcal{B}$ and with respect to $\widetilde{\mathcal{B}}$.

1.3. The Einstein Convention

1.3.1. A Convenient Summation Convention.

We start by setting a notation that will turn out to be useful later on. Recall that if $\mathcal{B} = \{b_1, b_2, b_3\}$ is a basis of a vector space V, any vector $v \in V$ can be written as

$$v = v^1 b_1 + v^2 b_2 + v^3 b_3 \tag{1.1}$$

for appropriate $v^1, v^2, v^3 \in \mathbb{R}$.

NOTATION. From now on, expressions like the one in (1.1) will be written as

$$v = \cancel{v^1 b_1 + v^2 b_2 + v^3 b_3} = v^j b_j \, . \tag{1.2}$$

That is, from now on when an index appears *twice – once as a subscript and once as a superscript* – in a term, we know that it means that there is a summation over all possible values of that index. The summation symbol will not be displayed.

On the other hand, indices that are not repeated in expressions like $a_{ij}x^k y^j$ are *free indices* not subject to summation.

EXAMPLES 1.24. For indices ranging over $\{1, 2, 3\}$, i.e. $n = 3$:

(1) The expression $a_{ij}x^i y^k$ means

$$a_{ij}x^i y^k = a_{1j}x^1 y^k + a_{2j}x^2 y^k + a_{3j}x^3 y^k \, ,$$

and could be called R^k_j (meaning that R^k_j and $a_{ij}x^i y^k$ both depend on the indices j and k).

(2) Likewise,

$$a_{ij}x^k y^j = a_{i1}x^k y^1 + a_{i2}x^k y^2 + a_{i3}x^k y^3 =: Q^k_i \, .$$

(3) Further

$$\begin{aligned} a_{ij}x^i y^j &= a_{11}x^1 y^1 + a_{12}x^1 y^2 + a_{13}x^1 y^3 \\ &\quad + a_{21}x^2 y^1 + a_{22}x^2 y^2 + a_{23}x^2 y^3 \\ &\quad + a_{31}x^3 y^1 + a_{32}x^3 y^2 + a_{33}x^3 y^3 =: P \end{aligned}$$

(4) An expression like

$$A^i B^j_{k\ell} C^\ell =: D^{ij}_k$$

makes sense. Here the indices i, j, k are free (i.e. free to range in $\{1, 2, \dots, n\}$) and ℓ is a summation index.

(5) On the other hand an expression like

$$E_{ij} F_\ell{}^{jk} G^\ell = H^{jk}_i$$

does *not* make sense because the expression on the left has only two free indices, i and k, while j and ℓ are summation indices and neither of them can appear on the right hand side.

NOTATION. Since $v^j b_j$ denotes a sum, we choose to denote the indices of the generic term of a sum with *capital letters*. For example, we write $v^I b_I$ and the above expressions could have been written as

(1)

$$a_{ij}x^i y^k = \sum_{I=1}^{3} a_{IJ} x^I y^K = a_{1j}x^1 y^k + a_{2j}x^2 y^k + a_{3j}x^3 y^k \,,$$

(2)

$$a_{ij}x^k y^j = \sum_{J=1}^{3} a_{IJ} x^K y^J = a_{i1}x^k y^1 + a_{i2}x^k y^2 + a_{i3}x^k y^3 \,.$$

(3)

$$\begin{aligned} a_{ij}x^i y^j &= \sum_{J=1}^{3}\sum_{I=1}^{3} a_{IJ} x^I y^J = \\ &= a_{11}x^1 y^1 + a_{12}x^1 y^2 + a_{13}x^1 y^3 \\ &+ a_{21}x^2 y^1 + a_{22}x^2 y^2 + a_{23}x^2 y^3 \\ &+ a_{31}x^3 y^1 + a_{32}x^3 y^2 + a_{33}x^3 y^3 \,. \end{aligned}$$

□

1.3.2. Change of Basis.

Let $\mathcal{B}$ and $\widetilde{\mathcal{B}}$ be two bases of a vector space V and let

$$L := L_{\widetilde{\mathcal{B}}\mathcal{B}} = \begin{bmatrix} L_1^1 & \dots & L_n^1 \\ \vdots & & \vdots \\ L_1^n & \dots & L_n^n \end{bmatrix} \tag{1.3}$$

be the **matrix of the change of basis** from the "old" basis $\mathcal{B}$ to the "new" basis $\widetilde{\mathcal{B}}$. Recall that the entries of the j-th column of L are the coordinates of the new basis vector $\tilde{b}_j$ with respect to the old basis $\mathcal{B}$.

Mnemonic: **Up**per indices go **up** to down, i.e., they are row indices.
Lower indices go **l**eft to right, i.e., they are column indices.

With the Einstein convention we can write

$$\boxed{\tilde{b}_j = L_j^i b_i}, \tag{1.4}$$

or, equivalently,

$$\boxed{\begin{pmatrix}\tilde{b}_1 & \dots & \tilde{b}_n\end{pmatrix} = \begin{pmatrix}b_1 & \dots & b_n\end{pmatrix} L},$$

where we use some convenient formal notation: The multiplication is to be performed with the usual rules for vectors and matrices, though, in this case, the entries of the row vectors $\begin{pmatrix}\tilde{b}_1 & \dots & \tilde{b}_n\end{pmatrix}$ and $\begin{pmatrix}b_1 & \dots & b_n\end{pmatrix}$ are not real numbers but vectors themselves.

If $\Lambda = L^{-1}$ denotes the matrix of the change of basis from $\widetilde{\mathcal{B}}$ to $\mathcal{B}$, then, using the same formal notation as above, we have

$$\boxed{\begin{pmatrix}b_1 & \dots & b_n\end{pmatrix} = \begin{pmatrix}\tilde{b}_1 & \dots & \tilde{b}_n\end{pmatrix}\Lambda}\,.$$

Equivalently, this can be written in compact form using the Einstein notation as

$$\boxed{b_j = \Lambda^i_j \tilde{b}_i}\,.$$

Analogously, the corresponding relations for the vector coordinates are

$$\begin{pmatrix} v^1 \\ \vdots \\ v^i \\ \vdots \\ v^n \end{pmatrix} = \begin{bmatrix} L^1_1 & \dots\dots & L^1_n \\ \vdots & & \vdots \\ L^i_1 & \dots\dots & L^i_n \\ \vdots & & \vdots \\ L^n_1 & \dots\dots & L^n_n \end{bmatrix} \begin{pmatrix} \tilde{v}^1 \\ \vdots \\ \vdots \\ \vdots \\ \tilde{v}^n \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \tilde{v}^1 \\ \vdots \\ \tilde{v}^i \\ \vdots \\ \tilde{v}^n \end{pmatrix} = \begin{bmatrix} \Lambda^1_1 & \dots\dots & \Lambda^1_n \\ \vdots & & \vdots \\ \Lambda^i_1 & \dots\dots & \Lambda^i_n \\ \vdots & & \vdots \\ \Lambda^n_1 & \dots\dots & \Lambda^n_n \end{bmatrix} \begin{pmatrix} v^1 \\ \vdots \\ \vdots \\ \vdots \\ v^n \end{pmatrix}$$

and these can be written with the Einstein convention respectively as

$$\boxed{v^i = L^i_j \tilde{v}^j} \quad \text{and} \quad \boxed{\tilde{v}^i = \Lambda^i_j v^j}\,, \tag{1.5}$$

or, in matrix notation,

$$[v]_{\mathcal{B}} = L_{\widetilde{\mathcal{B}}\mathcal{B}}[v]_{\widetilde{\mathcal{B}}} \quad \text{and} \quad [v]_{\widetilde{\mathcal{B}}} = (L_{\widetilde{\mathcal{B}}\mathcal{B}})^{-1}[v]_{\mathcal{B}} = L_{\mathcal{B}\widetilde{\mathcal{B}}}[v]_{\mathcal{B}}\,.$$

Important: Note how the coordinate vectors change in a way *opposite* to the basis change. Hence, we say that the coordinate vectors are **contravariant**[3] because they change by L^{-1} when the basis changes by L.

EXAMPLE 1.25. We consider the following two bases of $\mathbb{R}^2$

$$\begin{aligned} \mathcal{B} &= \Bigg\{ \underbrace{\begin{bmatrix}1\\0\end{bmatrix}}_{b_1}, \underbrace{\begin{bmatrix}2\\1\end{bmatrix}}_{b_2} \Bigg\} \\ \widetilde{\mathcal{B}} &= \Bigg\{ \underbrace{\begin{bmatrix}3\\1\end{bmatrix}}_{\tilde{b}_1}, \underbrace{\begin{bmatrix}-1\\-1\end{bmatrix}}_{\tilde{b}_2} \Bigg\} \end{aligned} \tag{1.6}$$

and we look for the matrix of the change of basis. Namely we look for a matrix L such that

$$\begin{bmatrix}3 & -1\\1 & -1\end{bmatrix} = \begin{pmatrix}\tilde{b}_1 & \tilde{b}_2\end{pmatrix} = \begin{pmatrix}b_1 & b_2\end{pmatrix} L = \begin{bmatrix}1 & 2\\0 & 1\end{bmatrix} L\,.$$

[3] In Latin *contra* means "contrary" or "against".

There are two alternative ways of finding L:

(1) *With matrix inversion:* Recall that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{D}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \tag{1.7}$$

where $D = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is the *determinant*. Thus

$$L = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

(2) *With Gauss-Jordan elimination:*

$$\left[\begin{array}{cc|cc} 1 & 2 & 3 & -1 \\ 0 & 1 & 1 & -1 \end{array}\right] \quad \leftrightsquigarrow \quad \left[\begin{array}{cc|cc} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{array}\right]$$

□

1.3.3. The Kronecker Delta Symbol.

NOTATION. The **Kronecker delta symbol** δ^i_j is defined as

$$\delta^i_j := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j\,. \end{cases} \tag{1.8}$$

EXAMPLES 1.26. If L is a matrix, the (i,j)-entry of L is the coefficient in the i-th row and j-th column, and is denoted by L^i_j.

(1) The $n \times n$ **identity matrix**

$$I = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix}$$

has (i,j)-entry equal to δ^i_j.

(2) Let L and M be two square matrices. The (i,j)-th entry of the product

$$ML = \begin{bmatrix} M^1_1 & \dots\dots & M^1_n \\ \vdots & & \vdots \\ M^i_1 & \dots\dots & M^i_n \\ \vdots & & \vdots \\ M^n_1 & \dots\dots & M^n_n \end{bmatrix} \begin{bmatrix} L^1_1 \dots & L^1_j & \dots L^1_n \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ L^n_1 \dots & L^n_j & \dots L^n_n \end{bmatrix}$$

equals the *dot product* of the i-th row of M and j-th column of L,

$$\begin{pmatrix} M^i_1 & \dots & M^i_n \end{pmatrix} \cdot \begin{pmatrix} L^1_j \\ \vdots \\ L^n_j \end{pmatrix} = M^i_1 L^1_j + \dots + M^i_n L^n_j\,,$$

or, using the Einstein convention,

$$M^i_k L^k_j \, .$$

Notice that, since in general $ML \neq LM$, it follows that

$$M^i_k L^k_j \neq L^i_k M^k_j = M^k_j L^i_k \, .$$

However, in the special case where $M = \Lambda = L^{-1}$, we have $\Lambda L = L\Lambda = I$ and here we can write

$$\Lambda^i_k L^k_j = \delta^i_j = L^i_k \Lambda^k_j \, .$$

□

REMARK 1.27. Using the Kronecker delta symbol we can check that the notations in (1.5) are all consistent. In fact, from (1.2) we should have

$$v^i b_i = v = \tilde{v}^i \tilde{b}_i \, , \tag{1.9}$$

and, in fact, using (1.5),

$$\tilde{v}^i \tilde{b}_i = \Lambda^i_j v^j L^k_i b_k = \delta^k_j v^j b_k = v^j b_j \, ,$$

where we used that $\Lambda^i_j L^k_i = \delta^k_j$ since $\Lambda = L^{-1}$.

Two words of warning:

- The two expressions $v^j b_j$ and $v^k b_k$ are identical, as the indices j and k are *dummy indices*, that is, can be replaced by other symbols throughout, without changing the meaning of the expression (as long as the symbols do not collide with other symbols already used in the expression).
- When multiplying two different expressions in Einstein notation, you should be careful to distinguish by different letters different summation indices. For example, if $\tilde{v}^i = \Lambda^i_j v^j$ and $\tilde{b}_i = L^j_i b_j$, in order to perform the multiplication $\tilde{v}^i \tilde{b}_i$ we have to make sure to replace one of the dummy indices in the two expressions. So, for example, we can write $\tilde{b}_i = L^k_i b_k$, so that $\tilde{v}^i \tilde{b}_i = \Lambda^i_j v^j L^k_i b_k$.

1.4. Linear Transformations

1.4.1. Linear Transformations as $(1,1)$-Tensors.

Recall that a linear transformation from V to itself, $T : V \to V$, is a function (or *map* or *transformation*) that satisfies the property

$$T(\alpha v + \beta w) = \alpha T(v) + \beta T(w) \, ,$$

for all $\alpha, \beta \in \mathbb{R}$ and all $v, w \in V$. Once we choose a basis $\mathcal{B} = \{b_1, \ldots b_n\}$ of V, the transformation T is represented by a matrix A, called the **matrix of the linear transformation with respect to that basis**. The columns of A are the coordinate vectors of $T(b_1), \ldots, T(b_n)$ with respect to $\mathcal{B}$. Then that matrix A gives the effect of

T on coordinate vectors as follows: If $T(v)$ is the value of the transformation T on the vector v, with respect to a basis $\mathcal{B}$ we have that

$$[v]_{\mathcal{B}} \longmapsto [T(v)]_{\mathcal{B}} = A[v]_{\mathcal{B}} . \tag{1.10}$$

If $\widetilde{\mathcal{B}}$ is another basis, we have also

$$[v]_{\widetilde{\mathcal{B}}} \longmapsto [T(v)]_{\widetilde{\mathcal{B}}} = \widetilde{A}[v]_{\widetilde{\mathcal{B}}} , \tag{1.11}$$

where now $\widetilde{A}$ is the matrix of the transformation T with respect to the basis $\widetilde{\mathcal{B}}$.

We want to find now the relation between A and $\widetilde{A}$. Let $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$ be the matrix of the change of basis from $\mathcal{B}$ to $\widetilde{\mathcal{B}}$. Then, for any $v \in V$,

$$[v]_{\widetilde{\mathcal{B}}} = L^{-1}[v]_{\mathcal{B}} . \tag{1.12}$$

In particular the above equation holds for the vector $T(v)$, that is

$$[T(v)]_{\widetilde{\mathcal{B}}} = L^{-1}[T(v)]_{\mathcal{B}} . \tag{1.13}$$

Using (1.12), (1.11), (1.13) and (1.10) in this order, we have

$$\widetilde{A}L^{-1}[v]_{\mathcal{B}} = \widetilde{A}[v]_{\widetilde{\mathcal{B}}} = [T(v)]_{\widetilde{\mathcal{B}}} = L^{-1}[T(v)]_{\mathcal{B}} = L^{-1}A[v]_{\mathcal{B}}$$

for every vector $v \in V$. If follows that $\widetilde{A}L^{-1} = L^{-1}A$ or equivalently

$$\widetilde{A} = L^{-1}AL , \tag{1.14}$$

which in Einstein notation reads

$$\widetilde{A}^i_j = \Lambda^i_k A^k_m L^m_j .$$

We say that the linear transformation T is a **tensor of type** $(1,1)$.

Example 1.28. Let $V = \mathbb{R}^2$ and let $\mathcal{B}$ and $\widetilde{\mathcal{B}}$ be the bases in Example 1.25. The matrices corresponding to the change of coordinates are

$$L := L_{\widetilde{\mathcal{B}}\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad \text{and} \qquad L^{-1} = \tfrac{1}{-2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} ,$$

where in the last equality we used the formula for the inverse of a matrix in (1.7).

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that in the basis $\mathcal{B}$ takes the form

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} .$$

Then according to (1.14) the matrix $\widetilde{A}$ of the linear transformation T with respect to the basis $\widetilde{\mathcal{B}}$ is

$$\widetilde{A} = L^{-1}AL = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 5 & -2 \\ -1 & 0 \end{bmatrix} .$$

□

EXAMPLE 1.29. We now look for the **standard matrix** of T, that is, the matrix M that represents T with respect to the standard basis of $\mathbb{R}^2$, which we denote by

$$\mathcal{E} := \Big\{ \underbrace{\begin{bmatrix}1\\0\end{bmatrix}}_{e_1}, \underbrace{\begin{bmatrix}0\\1\end{bmatrix}}_{e_2} \Big\}.$$

We want to apply again the formula (1.14) and hence we first need to find the matrix $S := L_{\mathcal{B}\mathcal{E}}$ of the change of basis from $\mathcal{E}$ to $\mathcal{B}$. Recall that the columns of S are the coordinates of b_j with respect to the basis $\mathcal{E}$, that is

$$S = \begin{bmatrix}1 & 2\\0 & 1\end{bmatrix}.$$

According to (1.14),

$$A = S^{-1}MS\,,$$

from which, using again (1.7), we obtain

$$M = SAS^{-1} = \begin{bmatrix}1 & 2\\0 & 1\end{bmatrix}\begin{bmatrix}1 & 3\\2 & 4\end{bmatrix}\begin{bmatrix}1 & -2\\0 & 1\end{bmatrix} = \begin{bmatrix}1 & 2\\0 & 1\end{bmatrix}\begin{bmatrix}1 & 1\\2 & 0\end{bmatrix} = \begin{bmatrix}5 & 1\\2 & 0\end{bmatrix}.$$

□

EXAMPLE 1.30. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the *orthogonal projection* onto the plane $\mathcal{P}$ of equation

$$2x + y - z = 0\,.$$

This means that the transformation T is characterized by the fact that

- it does not change vectors in the plane $\mathcal{P}$, and
- it sends vectors perpendicular to $\mathcal{P}$ to the zero vector in $\mathcal{P}$.

We want to find the standard matrix for T.

Idea: First compute the matrix of T with respect to a basis $\mathcal{B}$ of $\mathbb{R}^3$ well adapted to the problem, then use (1.14) after having found the matrix $L_{\mathcal{B}\mathcal{E}}$ of the change of basis.

To this purpose, we choose two linearly independent vectors in the plane $\mathcal{P}$ and a third vector perpendicular to $\mathcal{P}$. For instance, we set

$$\mathcal{B} := \Big\{ \underbrace{\begin{bmatrix}1\\0\\2\end{bmatrix}}_{b_1}, \underbrace{\begin{bmatrix}0\\1\\1\end{bmatrix}}_{b_2}, \underbrace{\begin{bmatrix}2\\1\\-1\end{bmatrix}}_{b_3} \Big\},$$

where the coordinates of b_1 and b_2 satisfy the equation of the plane, while the coordinates of b_3 are the coefficients of the equation describing $\mathcal{P}$. Let $\mathcal{E}$ be the standard basis of $\mathbb{R}^3$.

Since

$$T(b_1) = b_1, \qquad T(b_2) = b_2 \qquad \text{and } T(b_3) = 0\,,$$

the matrix of T with respect to $\mathcal{B}$ is simply

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{1.15}$$

where we recall that the j-th column is the coordinate vector $[T(b_j)]_{\mathcal{B}}$ of the vector $T(b_j)$ with respect to the basis $\mathcal{B}$.

The matrix of the change of basis from $\mathcal{E}$ to $\mathcal{B}$ is

$$L = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & -1 \end{bmatrix},$$

hence, by Gauss–Jordan elimination,

$$L^{-1} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{5}{6} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} & -\frac{1}{6} \end{bmatrix}.$$

Therefore

$$M = LAL^{-1} = \cdots = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{5}{6} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} & \frac{5}{6} \end{bmatrix}.$$

□

EXAMPLE 1.31. Let $V := \mathbb{R}[x]_2$ be the vector space of polynomials of degree ≤ 2, and let $T : \mathbb{R}[x]_2 \to \mathbb{R}[x]_2$ be the linear transformation given by differentiating a polynomial and then multiplying the derivative by x,

$$T(p(x)) := xp'(x),$$

so that $T(a + bx + cx^2) = x(b + 2cx) = bx + 2cx^2$. Let

$$\mathcal{B} := \{1, x, x^2\} \qquad \text{and} \qquad \widetilde{\mathcal{B}} := \{x, x-1, x^2-1\}$$

be two bases of $\mathbb{R}[x]_2$. Since

$$\begin{aligned} T(1) &= 0 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ T(x) &= x = 0 \cdot 1 + 1 \cdot x + 0 \cdot x^2 \\ T(x^2) &= 2x^2 = 0 \cdot 1 + 0 \cdot x + 2 \cdot x^2 \end{aligned}$$

and

$$\begin{aligned} T(x) &= x = 1 \cdot x + 0 \cdot (x-1) + 0 \cdot (x^2-1) \\ T(x-1) &= x = 1 \cdot x + 0 \cdot (x-1) + 0 \cdot (x^2-1) \\ T(x^2-1) &= 2x^2 = 2 \cdot x - 2 \cdot (x-1) + 2 \cdot (x^2-1), \end{aligned}$$

then

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \qquad \text{and} \qquad \widetilde{A} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & -2 \\ 0 & 0 & 2 \end{bmatrix}.$$

One can check that indeed $AL = L\widetilde{A}$ or, equivalently $\widetilde{A} = L^{-1}AL$, where

$$L = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is the matrix of the change of basis. □

1.4.2. Conjugate Matrices.

The above calculations can be summarized by the *commutativity* of the following diagram. Here, the vertical arrows correspond to the operation of change of basis from $\mathcal{B}$ to $\widetilde{\mathcal{B}}$ (recall that the coordinate vectors are contravariant tensors, that is, they transform as $[v]_{\widetilde{\mathcal{B}}} = L^{-1}[v]_{\mathcal{B}}$) and the horizontal arrows correspond to the operation of applying the transformation T with respect to the two different basis:

$$\begin{array}{ccc} [v]_{\mathcal{B}} & \xrightarrow{\;A\;} & [T(v)]_{\mathcal{B}} \\ {\scriptstyle L^{-1}}\big\downarrow & & \big\downarrow{\scriptstyle L^{-1}} \\ [v]_{\widetilde{\mathcal{B}}} & \xrightarrow[\;\widetilde{A}\;]{} & [T(v)]_{\widetilde{\mathcal{B}}} \,. \end{array}$$

Saying that *the diagram is commutative* is saying that if one starts from the upper left hand corner, reaching the lower right hand corner following either one of the two paths has exactly the same effect. In other words, changing coordinates first then applying the transformation T yields exactly the same affect as applying first the transformation T and then the change of coordinates, that is, $L^{-1}A = \widetilde{A}L^{-1}$ or, equivalently,

$$\widetilde{A} = L^{-1}AL\,.$$

In this case we say that A and $\widetilde{A}$ are *conjugate* matrices. This means that A and $\widetilde{A}$ represent the same transformation with respect to different bases.

Definition 1.32. We say that two matrices A and $\widetilde{A}$ are **conjugate** if there exists and invertible matrix L such that $\widetilde{A} = L^{-1}AL$.

Example 1.33. The three matrices from Example 1.28 and Example 1.29

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \qquad M = \begin{bmatrix} 5 & 1 \\ 2 & 0 \end{bmatrix} \qquad \text{and} \qquad \widetilde{A} = \begin{bmatrix} 5 & -2 \\ -1 & 0 \end{bmatrix}$$

are all conjugate. Indeed, we have

$$\widetilde{A} = L^{-1}AL\,, \qquad A = S^{-1}MS \quad \text{and} \quad \widetilde{A} = R^{-1}MR\,,$$

where L and S may be found in those examples and where $R := SL$. □

We now review some facts about conjugate matrices. Recall that the **characteristic polynomial** of a square matrix A is the polynomial

$$p_A(\lambda) := \det(A - \lambda I)\,.$$

Let us assume that A and $\widetilde{A}$ are conjugate matrices, that is $\widetilde{A} = L^{-1}AL$ for some invertible matrix L. Then

$$\begin{aligned} p_{\widetilde{A}}(\lambda) &= \det(\widetilde{A} - \lambda I) = \det(L^{-1}AL - \lambda L^{-1}IL) \\ &= \det(L^{-1}(A - \lambda I)L) \\ &= \cancel{(\det L^{-1})}\det(A - \lambda I)\cancel{(\det L)} = p_A(\lambda)\,, \end{aligned} \tag{1.16}$$

which means that any two conjugate matrices have the same characteristic polynomial.

Recall that the **eigenvalues** of a matrix A are the roots of its characteristic polynomial and we here usually allow complex roots. Then, by the so-called *fundamental theorem of Algebra*, each $n \times n$ matrix has n (real or complex) eigenvalues counted with multiplicities as polynomial roots. Recall also the definitions of *determinant* and *trace* of a square matrix. By analysing the characteristic polynomial, we see that

(1) the **determinant** of a matrix is equal to the product of its eigenvalues (multiplied with multiplicities), and
(2) the **trace** of a matrix is equal to the sum of its eigenvalues (added with multiplicities).

From (1.16) if follows that, if the matrices A and $\widetilde{A}$ are conjugate, then:

- A and $\widetilde{A}$ have the same size;
- the eigenvalues of A (as well as their multiplicities) are the same as those of $\widetilde{A}$;
- $\det A = \det \widetilde{A}$;
- $\operatorname{tr} A = \operatorname{tr} \widetilde{A}$;
- A is invertible if and only if $\widetilde{A}$ is invertible.

Example 1.34. The matrices $A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ and $A' = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ are not conjugate. In fact, A is invertible, as $\det A = -2 \neq 0$, while $\det A' = 0$, so that A' is not invertible. □

1.4.3. Eigenbases.

The possibility of choosing different bases is very important and often simplifies the calculations. Example 1.30 is such an example, where we choose an appropriate basis according to the specific problem. Other times, a basis can be chosen according to the symmetries and, completely at the opposite side, sometime there is just not a basis that is a preferred one. In the context of a linear transformation $T : V \to V$, a basis that is particularly convenient, when it exists, is an **eigenbasis** for that linear transformation.

Recall that an **eigenvector** of a linear transformation $T : V \to V$ is a vector $v \neq 0$ such that $T(v)$ is a multiple of v, say $T(v) = \lambda v$ and, in that case, the scaling number λ is called an **eigenvalue** of T.

An **eigenbasis** is a basis of V consisting of eigenvectors of a linear transformation $T : V \to V$. The point of having an eigenbasis is that, with respect to this eigenbasis, the linear transformation is representable by a *diagonal* matrix, D. Hence, the initial matrix representative A is actually conjugate to a diagonal matrix D. A linear transformation $T : V \to V$ for which an eigenbasis exists is then called **diagonalizable**.[4]

Given a linear transformation $T : V \to V$, in order to find an eigenbasis of T, we first represent T by some matrix A (with respect to some chosen basis of V), and then perform the following steps:

(1) We find the eigenvalues by determining the roots of the characteristic polynomial of A (often allowing complex roots).
(2) For each eigenvalue λ, we find the corresponding eigenvectors by looking for the nonzero vectors in the so-called **eigenspace**

$$E_\lambda := \ker(A - \lambda I) .$$

When considering complex eigenvalues, the eigenspaces are determined as subspaces of the *complex vector space* $\mathbb{C}^n$. However, in this text, we concentrate on real cases.
(3) We determine whether there exists an eigenbasis.

We will illustrate this in the following examples.

EXAMPLE 1.35. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by the matrix $A = \begin{bmatrix} 3 & -4 \\ -4 & -3 \end{bmatrix}$ with respect to the standard basis of $\mathbb{R}^2$.

(1) The eigenvalues are the roots of the characteristic polynomial $p_\lambda(A)$. Since

$$\begin{aligned} p_A(\lambda) = \det(A - \lambda I) &= \det \begin{bmatrix} 3-\lambda & -4 \\ -4 & -3-\lambda \end{bmatrix} \\ &= (3-\lambda)(-3-\lambda) - 16 = \lambda^2 - 25 = (\lambda - 5)(\lambda + 5) , \end{aligned}$$

hence $\lambda = \pm 5$ are the eigenvalues of A.
(2) If λ is an eigenvalue of A, the eigenspace corresponding to λ is given by $E_\lambda = \ker(A - \lambda I)$. Note that

$$v \in E_\lambda \quad \Longleftrightarrow \quad Av = \lambda v .$$

[4]In general, when an eigenbasis does not exist, it is still possible to find a basis, with respect to which the linear transformation is as simple as possible, i.e., as close as possible to being diagonal. Such a best matrix representative of $T : V \to V$ is called a **Jordan canonical form** and is, of course, conjugate to the first matrix representative A. In this text, we will not address such more general canonical forms.

With our choice of A and with the resulting eigenvalues, we have

$$E_5 = \ker(A - 5I) = \ker \begin{bmatrix} -2 & -4 \\ -4 & -8 \end{bmatrix} = \operatorname{span} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$
$$E_{-5} = \ker(A + 5I) = \ker \begin{bmatrix} 8 & -4 \\ -4 & 2 \end{bmatrix} = \operatorname{span} \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

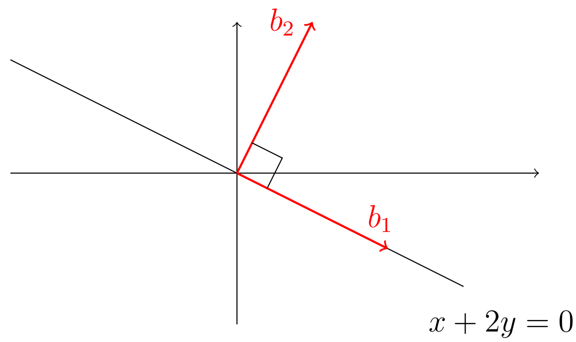
(3) The following is an eigenbasis for this linear transformation:

$$\widetilde{\mathcal{B}} = \left\{ \tilde{b}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \tilde{b}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

and

$$T(\tilde{b}_1) = 5\tilde{b}_1 = 5 \cdot \tilde{b}_1 + 0 \cdot \tilde{b}_2$$
$$T(\tilde{b}_2) = -5\tilde{b}_2 = 0 \cdot \tilde{b}_1 - 5 \cdot \tilde{b}_2 \,,$$

so that $\tilde{A} = \begin{bmatrix} 5 & 0 \\ 0 & -5 \end{bmatrix}$.



Notice that the eigenspace E_5 consists of vectors on the line $x + 2y = 0$ and these vectors get scaled by the transformation T by a factor of 5. On the other hand, the eigenspace E_{-5} consists of vectors perpendicular to the line $x+2y = 0$ and these vectors get flipped by the transformation T and then also scaled by a factor of 5. Hence T is just the reflection across the line $x+2y = 0$ followed by multiplication by 5.

□

Example 1.36. Now let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by the matrix $A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$ with respect to the standard basis of $\mathbb{R}^2$.

(1) The eigenvalues are the roots of the characteristic polynomial:

$$p_A(\lambda) = \det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & 2 \\ 4 & 3-\lambda \end{bmatrix}$$
$$= (1-\lambda)(3-\lambda) - 2 \cdot 4 = \lambda^2 - 4\lambda - 5 = (\lambda+1)(\lambda-5)\,,$$

hence $\lambda = -1$ and $\lambda = 5$ are the eigenvalues of A.

(2) If λ is an eigenvalue of A, the eigenspace corresponding to λ is given by $E_\lambda = \ker(A - \lambda I)$. In this case we have

$$E_{-1} = \ker(A + I) = \ker \begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix} = \operatorname{span} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
$$E_5 = \ker(A - 5I) = \ker \begin{bmatrix} -4 & 2 \\ 4 & -2 \end{bmatrix} = \operatorname{span} \begin{bmatrix} 1 \\ 2 \end{bmatrix} .$$

(3) The following is an eigenbasis for this linear transformation:

$$\widetilde{\mathcal{B}} = \left\{ \tilde{b}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \tilde{b}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\} .$$

We have $T(\tilde{b}_1) = -\tilde{b}_1$ and $T(\tilde{b}_2) = 5\tilde{b}_2$, hence $\tilde{A} = \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix}$.

□

Example 1.37. Now let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by the matrix $A = \begin{bmatrix} 5 & -3 \\ 3 & -1 \end{bmatrix}$ with respect to the standard basis of $\mathbb{R}^2$.

(1) The eigenvalues are the roots of the characteristic polynomial:

$$\begin{aligned} p_A(\lambda) = \det(A - \lambda I) &= \det \begin{bmatrix} 5 - \lambda & -3 \\ 3 & -1 - \lambda \end{bmatrix} \\ &= (5 - \lambda)(-1 - \lambda) + 9 = \lambda^2 - 4\lambda + 4 = (\lambda - 2)^2 , \end{aligned}$$

hence $\lambda = 2$ is the only eigenvalue of A.

(2) The eigenspace corresponding to $\lambda = 2$ is

$$E_2 = \ker(A - 2I) = \ker \begin{bmatrix} 3 & -3 \\ 3 & -3 \end{bmatrix} = \operatorname{span} \begin{bmatrix} 1 \\ 1 \end{bmatrix} .$$

(3) Since we cannot find two linearly independent eigenvectors (in order to form a basis of $\mathbb{R}^2$), we conclude that in this case there is *no* eigenbasis for this linear transformation.

□

Summarizing, in Examples 1.28 and 1.29, we looked at how the matrix of a transformation changes with respect to two different bases that we were given. In Example 1.30, we looked for a particular basis appropriate to the transformation at hand. In Example 1.35, we looked for an eigenbasis with respect to the given transformation. Example 1.30 in this respect fits into the same framework as Example 1.35, but the orthogonal projection has a zero eigenvalue (see (1.15)). Example 1.36 illustrates how eigenvectors, in general, need not be orthogonal. In Example 1.37 we see that sometimes an eigenbasis does not exist.

CHAPTER 2

Multilinear Forms

2.1. Linear Forms

Linear forms on a vector space V are defined as linear real-valued functions on V. We will see that linear forms behave very much like vectors, only that they are elements *not* of V, but of a different, yet related, vector space. Whereas we represent regular vectors (from V) by column vectors once a basis is fixed, we will represent linear forms (on V) by row vectors. Then the value of a linear form on a specific vector is simply given by by the matrix product with the row vector (linear form) on the left and the column vector (actual vector) on the right.

2.1.1. Definition and Examples.

DEFINITION 2.1. Let V be a vector space. A **linear form** on V is a map $\alpha : V \to \mathbb{R}$ such that for every $a, b \in \mathbb{R}$ and for every $v, w \in V$

$$\alpha(av + bw) = a\alpha(v) + b\alpha(w) .$$

Alternative terminologies for "linear form" are **tensor of type** $(0, 1)$, **1-form**, **linear functional** and **covector**.

EXERCISE 2.2. If $V = \mathbb{R}^3$, which of the following are linear forms?

(1) $\alpha(x, y, z) := xy + z$;
(2) $\alpha(x, y, z) := x + y + z + 1$;
(3) $\alpha(x, y, z) := \pi x - \frac{7}{2} z$.

EXERCISE 2.3. If V is the infinite dimensional vector space of continuous functions $f : \mathbb{R} \to \mathbb{R}$, which of the following are linear forms?

(1) $\alpha(f) := f(7) - f(0)$;
(2) $\alpha(f) := \int_0^{33} e^x f(x) dx$;
(3) $\alpha(f) := e^{f(4)}$.

EXAMPLE 2.4. [Coordinate forms] This is a most important example of linear form. Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis of V and let $v = v^i b_i \in V$ be a generic vector. Define $\beta^i : V \to \mathbb{R}$ by

$$\boxed{\beta^i(v) := v^i} , \tag{2.1}$$

that is β^i will extract the i-th coordinate of a vector with respect to the basis $\mathcal{B}$. The linear form β^i is called **coordinate form**. Notice that

$$\beta^i(b_j) = \delta^i_j\,, \tag{2.2}$$

since the i-th coordinate of the basis vector b_j with respect to the basis $\mathcal{B}$ is equal to 1 if $i = j$ and 0 otherwise. $\square$

EXAMPLE 2.5. Let $V = \mathbb{R}^3$ and let $\mathcal{E}$ be its standard basis. The three coordinate forms are defined by

$$\beta^1 \begin{bmatrix} x \\ y \\ z \end{bmatrix} := x\,, \quad \beta^2 \begin{bmatrix} x \\ y \\ z \end{bmatrix} := y\,, \quad \beta^3 \begin{bmatrix} x \\ y \\ z \end{bmatrix} := z\,.$$

$\square$

EXAMPLE 2.6. Let $V = \mathbb{R}^2$ and let $\mathcal{B} := \Bigg\{ \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{b_1}, \underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{b_2} \Bigg\}$. We want to describe the elements of $\mathcal{B}^* := \{\beta^1, \beta^2\}$, in other words we want to find

$$\beta^1(v) \qquad \text{and} \qquad \beta^2(v)$$

for a generic vector $v \in V$.

To this purpose we need to find $[v]_{\mathcal{B}}$. Recall that if $\mathcal{E}$ denotes the standard basis of $\mathbb{R}^2$ and $L := L_{\mathcal{B}\mathcal{E}}$ the matrix of the change of coordinate from $\mathcal{E}$ to $\mathcal{B}$, then

$$[v]_{\mathcal{B}} = L^{-1}[v]_{\mathcal{E}} = L^{-1} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} .$$

Since

$$L = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and hence

$$L^{-1} = \tfrac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} ,$$

then

$$[v]_{\mathcal{B}} = \begin{pmatrix} \frac{1}{2}(v^1 + v^2) \\ \frac{1}{2}(v^1 - v^2) \end{pmatrix} .$$

Thus, according to the definition (2.1), we deduce that

$$\beta^1(v) = \tfrac{1}{2}(v^1 + v^2) \quad \text{and} \quad \beta^2(v) = \tfrac{1}{2}(v^1 - v^2)\,.$$

$\square$

2.1.2. Dual Space and Dual Basis.

We define

$$V^* := \{\text{all linear forms } \alpha : V \to \mathbb{R}\}\,,$$

and call this the **dual** (or *dual space*) of V.

EXERCISE 2.7. Check that V^* is a vector space whose null vector is the linear form identically equal to zero.

REMARK 2.8. Just like any function, two linear forms on V are equal if and only if their values are the same when applied to *each* vector in V. However, because of the defining properties of linear forms, to determine whether two linear forms are equal, it is *enough to check that they are equal on each element of a basis of* V. In fact, let $\alpha, \alpha' \in V^*$, let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis of V and suppose that we know that

$$\alpha(b_j) = \alpha'(b_j)$$

for all $1 \le j \le n$. We verify that this implies that they are the same when applied to each vector $v \in V$. In fact let $v = v^j b_j$ its representation with respect to the basis $\mathcal{B}$. Then we have

$$\alpha(v) = \alpha(v^j b_j) = v^j \alpha(b_j) = v^j \alpha'(b_j) = \alpha'(v^j b_j) = \alpha'(v)\,.$$

□

PROPOSITION 2.9. *Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis of V and $\beta^1, \dots, \beta^n$ the corresponding coordinate forms. Then $\mathcal{B}^* := \{\beta^1, \dots, \beta^n\}$ is a basis of V^*. As a consequence*

$$\dim V = \dim V^*\,.$$

PROOF. According to Definition 1.17, we need to check that the linear forms in $\mathcal{B}^*$

(1) are linearly independent and
(2) span V^*.

(1) We need to check that the only linear combination of $\beta^1, \dots, \beta^n$ that yields the zero linear form is the trivial linear combination. Let $c_i\beta^i = 0$ be a linear combination of the β^i. Then for every basis vector b_j, with $j = 1, \dots, n$,

$$0 = (c_i\beta^i)(b_j) = c_i(\beta^i(b_j)) = c_i\delta^i_j = c_j\,,$$

thus showing the linear independence.

(2) To check that $\mathcal{B}^*$ spans V we need to verify that any $\alpha \in V^*$ is a linear combination of $\beta^1, \dots, \beta^n$, that is that we can find $\alpha_i \in \mathbb{R}$ such that

$$\alpha = \alpha_i\beta^i \tag{2.3}$$

To find such α_i we apply both sides of (2.3) to the j-th basis vector b_i, and we obtain

$$\alpha(b_j) = \alpha_i\beta^i(b_j) = \alpha_i\delta^i_j = \alpha_j\,, \tag{2.4}$$

which identifies the coefficients in (2.3).

By hypothesis α is a linear form and, since V^* is a vector space, also $\alpha(b_i)\beta^i$ is a linear form. Moreover, we have just verified that these two linear form coincide on the basis vectors. By Remark 2.8 the two linear forms are the same and, hence, we have written α as a linear combination of the coordinate forms. This completes the proof that the coordinate forms form a basis of the dual. □

The basis $\mathcal{B}^*$ of V^* is called the **basis of V^* dual to $\mathcal{B}$**. We emphasize that the components (or coordinates) of a linear form α with respect to $\mathcal{B}^*$ are exactly the values of α on the elements of $\mathcal{B}$, as we found in the above proof:

$$\boxed{\alpha_i = \alpha(b_i)}\,.$$

We build with these the coordinate-vector of α as a *row-vector*:

$$[\alpha]_{\mathcal{B}^*} := \begin{pmatrix}\alpha_1 & \dots & \alpha_n\end{pmatrix} .$$

Example 2.10. Let $V = \mathbb{R}[x]_2$ be the vector space of polynomials of degree ≤ 2, let $\alpha : V \to \mathbb{R}$ be the linear form given by

$$\alpha(p) := p(2) - p'(2) \tag{2.5}$$

and let $\mathcal{B}$ be the basis $\{1, x, x^2\}$ of V. In this example, we want to:

(1) find the components of α with respect to $\mathcal{B}^*$;
(2) describe the basis $\mathcal{B}^* = \{\beta^1, \beta^2, \beta^3\}$;

(1) Since

$$\begin{aligned}\alpha_1 &= \alpha(b_1) = \alpha(1) = 1 - 0 = 1\\ \alpha_2 &= \alpha(b_2) = \alpha(x) = 2 - 1 = 1\\ \alpha_3 &= \alpha(b_3) = \alpha(x^2) = 4 - 4 = 0\,,\end{aligned}$$

then

$$[\alpha]_{\mathcal{B}^*} = \begin{pmatrix}1 & 1 & 0\end{pmatrix} . \tag{2.6}$$

(2) The generic element $p(x) \in \mathbb{R}[x]_2$ written as combination of basis elements $1, x$ and x^2 is

$$p(x) = a + bx + cx^2\,.$$

Hence $\mathcal{B}^* = \{\beta^1, \beta^2, \beta^3\}$, is given by

$$\begin{aligned}\beta^1(a + bx + cx^2) &= a\\ \beta^2(a + bx + cx^2) &= b\\ \beta^3(a + bx + cx^2) &= c\,.\end{aligned} \tag{2.7}$$

□

REMARK 2.11. Note that we have to be careful when referring to a "dual basis" of V^*, as for every basis $\mathcal{B}$ of V there is going to be a basis $\mathcal{B}^*$ of V^* dual to the basis $\mathcal{B}$. In the next section we are going to see how a dual basis transforms with a change of basis.

2.1.3. Covariance of Linear Forms.

We want to examine how a linear form $\alpha : V \to \mathbb{R}$ behaves with respect to a change a basis in V. To this purpose, let

$$\mathcal{B} = \{b_1, \dots, b_n\} \quad \text{and} \quad \widetilde{\mathcal{B}} := \{\tilde{b}_1, \dots, \tilde{b}_n\}$$

be two bases of V and let

$$\mathcal{B}^* := \{\beta^1, \dots, \beta^n\} \quad \text{and} \quad \widetilde{\mathcal{B}}^* := \{\tilde{\beta}^1, \dots, \tilde{\beta}^n\}$$

be the corresponding dual bases. Let

$$[\alpha]_{\mathcal{B}^*} = \begin{pmatrix}\alpha_1 & \dots & \alpha_n\end{pmatrix} \quad \text{and} \quad [\alpha]_{\widetilde{\mathcal{B}}^*} = \begin{pmatrix}\widetilde{\alpha}_1 & \dots & \widetilde{\alpha}_n\end{pmatrix}$$

be the coordinate vectors of α with respect to $\mathcal{B}^*$ and $\widetilde{\mathcal{B}}^*$, that is

$$\alpha(b_i) = \alpha_i \quad \text{and} \quad \alpha(\tilde{b}_i) = \widetilde{\alpha}_i\,.$$

Let $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$ be the matrix of the change of basis in (1.3)

$$\tilde{b}_j = L^i_j b_i\,.$$

Then

$$\widetilde{\alpha}_j = \alpha(\tilde{b}_j) = \alpha(L^i_j b_i) = L^i_j \alpha(b_i) = L^i_j \alpha_i = \alpha_i L^i_j\,, \tag{2.8}$$

so that

$$\boxed{\widetilde{\alpha}_j = \alpha_i L^i_j}\,. \tag{2.9}$$

EXERCISE 2.12. Verify that (2.9) is equivalent to saying that

$$[\alpha]_{\widetilde{\mathcal{B}}^*} = [\alpha]_{\mathcal{B}^*} L\,. \tag{2.10}$$

Note that we have exchanged the order of α_i and L^i_j in the last equation in (2.8) to respect the order in which the matrix multiplication in (2.10) has to be performed. This was possible because both α_i and L^i_j are real numbers.

We say that a linear form α is **covariant** because its components change by L when the basis changes by L.[5] A linear form α is hence a **covariant tensor** or a **tensor of type** $(0, 1)$.

[5] In Latin, the prefix *co* means "joint".

EXAMPLE 2.13. We continue with Example 2.10. We consider the bases as in Example 1.31, that is

$$\mathcal{B} := \{1, x, x^2\} \qquad \text{and} \qquad \widetilde{\mathcal{B}} := \{x, x-1, x^2-1\}$$

and the linear form $\alpha : V \to \mathbb{R}$ as in (2.5). We will:

(1) find the components of α with respect to $\mathcal{B}^*$;
(2) describe the basis $\mathcal{B}^* = \{\beta^1, \beta^2, \beta^3\}$;
(3) find the components of α with respect to $\widetilde{\mathcal{B}}^*$;
(4) describe the basis $\widetilde{\mathcal{B}}^* = \{\tilde\beta^1, \tilde\beta^2, \tilde\beta^3\}$;
(5) find the matrix of change of basis $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$ and compute $\Lambda = L^{-1}$;
(6) check the covariance of α;
(7) check the contravariance of $\mathcal{B}^*$.

(1) This is done in (2.6).

(2) This is done in (2.7).

(3) We proceed as in (2.6). Namely,

$$\begin{aligned}
\widetilde\alpha_1 &= \alpha(\tilde b_1) = \alpha(x) = 2 - 1 = 1 \\
\widetilde\alpha_2 &= \alpha(\tilde b_2) = \alpha(x-1) = 1 - 1 = 0 \\
\widetilde\alpha_3 &= \alpha(\tilde b_3) = \alpha(x^2-1) = 3 - 4 = -1 \,,
\end{aligned}$$

so that

$$[\alpha]_{\widetilde{\mathcal{B}}^*} = \begin{pmatrix} 1 & 0 & -1 \end{pmatrix} .$$

(4) Since $\tilde\beta^i(v) = \tilde v^i$, to proceed as in (2.7) we first need to write the generic polynomial $p(x) = a + bx + cx^2$ as a linear combination of elements in $\widetilde{\mathcal{B}}$, namely we need to find $\tilde a, \tilde b$ and $\tilde c$ such that

$$p(x) = a + bx + cx^2 = \tilde a x + \tilde b(x-1) + \tilde c(x^2-1) \,.$$

By multiplying and collecting the terms, we obtain that

$$\begin{cases} -\tilde b - \tilde c = a \\ \tilde a + \tilde b = b \\ \tilde c = c \end{cases} \qquad \text{that is} \qquad \begin{cases} \tilde a = a + b + c \\ \tilde b = -a - c \\ \tilde c = c \,. \end{cases}$$

Hence

$$p(x) = a + bx + cx^2 = (a+b+c)x + (-a-c)(x-1) + c(x^2-1) \,,$$

so that it follows that

$$\begin{aligned}
\tilde\beta^1(p(x)) &= a + b + c \\
\tilde\beta^2(p(x)) &= -a - c \\
\tilde\beta^3(p(x)) &= c \,,
\end{aligned}$$

(5) The matrix of the change of basis is given by

$$L := L_{\widetilde{\mathcal{B}}\mathcal{B}} = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} ,$$

since for example $\tilde{b}_3$ can be written as a linear combination with respect to $\mathcal{B}$ as $\tilde{b}_3 = x^2 - 1 = -1b_1 + 0b_2 + 1b_3$, and these coordinates $-1, 0, 1$ build the third column of L.

To compute $\Lambda = L^{-1}$ we can use the Gauss–Jordan elimination process

$$\left[\begin{array}{ccc|ccc} 0 & -1 & -1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \quad \leftrightsquigarrow \quad \dots \quad \leftrightsquigarrow \quad \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

Therefore, we have

$$\Lambda = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} .$$

(6) The linear form α is indeed *covariant*, since

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \end{pmatrix} L = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{pmatrix} 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} \widetilde{\alpha}_1 & \widetilde{\alpha}_2 & \widetilde{\alpha}_3 \end{pmatrix} .$$

(7) The dual basis $\mathcal{B}^*$ is *contravariant*, since

$$\begin{pmatrix} \tilde{\beta}^1 \\ \tilde{\beta}^2 \\ \tilde{\beta}^3 \end{pmatrix} = \Lambda \begin{pmatrix} \beta^1 \\ \beta^2 \\ \beta^3 \end{pmatrix} ,$$

as it can be verified by evaluating both sides on an arbitrary vector $p(x) = a + bx + cx^2$:

$$\Lambda \begin{pmatrix} \beta^1(p) \\ \beta^2(p) \\ \beta^3(p) \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a + b + c \\ -a - c \\ c \end{pmatrix} = \begin{pmatrix} \tilde{\beta}^1(p) \\ \tilde{\beta}^2(p) \\ \tilde{\beta}^3(p) \end{pmatrix} .$$

□

2.1.4. Contravariance of Dual Bases.

In fact, statement (7) in Example 2.10 holds in general, namely:

PROPOSITION 2.14. *Dual bases are* **contravariant**.

PROOF. We will check that when bases $\mathcal{B}$ and $\widetilde{\mathcal{B}}$ are related by

$$\tilde{b}_j = L^i_j b_i$$

V real vector space with $\dim V = n$ $\mathcal{B} = \{b_1, \dots, b_n\}$ basis of V $\widetilde{\mathcal{B}} := \{\tilde{b}_1, \dots, \tilde{b}_n\}$ another basis of V $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$ =matrix of the change of basis from $\mathcal{B}$ to $\widetilde{\mathcal{B}}$ Then we have $\tilde{b}_j = L^i_j b_i$ or $\begin{pmatrix}\tilde{b}_1 & \dots & \tilde{b}_n\end{pmatrix} = \begin{pmatrix}b_1 & \dots & b_n\end{pmatrix} L$ covariance of a basis	$V^* = \{\text{linear forms } \alpha : V \to \mathbb{R}\}$ dual vector space to V $\mathcal{B}^* = \{\beta^1, \dots, \beta^n\}$ dual basis of V^* w.r.t. $\mathcal{B}$ $\widetilde{\mathcal{B}}^* = \{\tilde{\beta}^1, \dots, \tilde{\beta}^n\}$ dual basis of V^* w.r.t. $\widetilde{\mathcal{B}}^*$ $\Lambda = L^{-1}$ =matrix of the change of basis from $\widetilde{\mathcal{B}}$ to $\mathcal{B}$ Then we have $\tilde{\beta}^i = \Lambda^i_j \beta^j$ or $\begin{pmatrix}\tilde{\beta}^1 \\ \vdots \\ \tilde{\beta}^n\end{pmatrix} = L^{-1} \begin{pmatrix}\beta^1 \\ \vdots \\ \beta^n\end{pmatrix}$ contravariance of the dual basis
If v is any vector in V, then $v = v^i b_i = \tilde{v}^i \tilde{b}_i$, where $\tilde{v}^i = \Lambda^i_j v^j$ i.e. $[v]_{\widetilde{\mathcal{B}}} = L^{-1}[v]_{\mathcal{B}}$ or $\begin{pmatrix}\tilde{v}^1 \\ \vdots \\ \tilde{v}^n\end{pmatrix} = L^{-1} \begin{pmatrix}v^1 \\ \vdots \\ v^n\end{pmatrix}$ contravariance of the coordinate vectors $\rightarrow$ vectors are $(1, 0)$-tensors	If α is any linear form in V^*, then $\alpha = \alpha_j \beta^j = \tilde{\alpha}_j \tilde{\beta}^j$, where $\tilde{\alpha}_j = L^i_j \alpha_i$ i.e. $[\alpha]_{\widetilde{\mathcal{B}}^*} = [\alpha]_{\mathcal{B}^*} L$ or $\begin{pmatrix}\tilde{\alpha}_1 & \dots & \tilde{\alpha}_n\end{pmatrix} = \begin{pmatrix}\alpha_1 & \dots & \alpha_n\end{pmatrix} L$ covariance of linear forms components $\rightarrow$ linear forms are $(0, 1)$-tensors

TABLE 1. Summary regarding duality

the corresponding dual bases $\mathcal{B}^*$ and $\widetilde{\mathcal{B}}^*$ of V^* are related by

$$\boxed{\tilde{\beta}^j = \Lambda^j_i \beta^i}. \tag{2.11}$$

It is enough to check that the $\Lambda^j_i \beta^i$ are *dual* to the $\tilde{b}_j$. In fact, since $\Lambda L = I$, then

$$(\Lambda^k_\ell \beta^\ell)(\tilde{b}_j) = (\Lambda^k_\ell \beta^\ell)(L^i_j b_i) = \Lambda^k_\ell L^i_j \beta^\ell(b_i) = \Lambda^k_\ell L^i_j \delta^\ell_i = \Lambda^k_i L^i_j = \delta^k_j = \beta^j(\tilde{b}_j)\,.$$

□

In Table 1, you can find a summary of the properties that bases and dual bases, coordinate vectors and components of linear forms satisfy with respect to a change of basis and hence whether they are covariant or contravariant. Moreover, Table 2 summarizes the characteristics of covariance and contravariance.

	covariance of a tensor	contravariace of a tensor
is denoted by	lower indices	upper indices
coordinate-vectors are indicated as	row vectors	column vectors
the tensor transforms w.r.t. a change of basis from $\mathcal{B}$ to $\widetilde{\mathcal{B}}$ by multiplication with	L on the right	L^{-1} on the left
(for later use) a tensor of type (p, q) has	covariant order q	contravariant order p

TABLE 2. Covariance vs. contravariance

2.2. Bilinear Forms

2.2.1. Definition and Examples.

DEFINITION 2.15. A **bilinear form** on V is a function $\varphi : V \times V \to \mathbb{R}$ that is linear in each variable, that is

$$\varphi(u, \lambda v + \mu w) = \lambda\varphi(u, v) + \mu\varphi(u, w)$$
$$\varphi(\lambda v + \mu w, u) = \lambda\varphi(v, u) + \mu\varphi(w, u)\,,$$

for every $\lambda, \mu \in \mathbb{R}$ and for every $u, v, w \in V$.

EXAMPLES 2.16. Let $V = \mathbb{R}^n$.

(1) If $v, w \in \mathbb{R}^n$, the **dot product** (or **scalar product**) defined as

$$\varphi(v, w) := v \cdot w = |v|\,|w| \cos\theta\,,$$

where θ is the angle between v and w is a bilinear form.

(2) Let $n = 3$. Choose a vector $u \in \mathbb{R}^3$ and for any two vectors $v, w \in \mathbb{R}^3$, denote by $v \times w$ their **cross product**. The **scalar triple product**

$$\varphi_u(v, w) := u \cdot (v \times w) = \det \begin{bmatrix} u \\ v \\ w \end{bmatrix} \tag{2.12}$$

is a bilinear form in v and w, where $\begin{bmatrix} u \\ v \\ w \end{bmatrix}$ denotes the matrix with rows u, v and w. The quantity $\varphi_u(v, w)$ calculates the signed volume of the parallelepiped spanned by u, v, w: the sign of $\varphi_u(v, w)$ depends on the orientation of the triple u, v, w.

Since the cross product is defined *only* in $\mathbb{R}^3$, in contrast with the scalar product, the scalar triple product cannot be defined in $\mathbb{R}^n$ with $n \neq 3$ (though there is a formula for an n dimensional paralleledipied involving some "generalization" of it).

□

EXERCISE 2.17. Verify the equality in (2.12) using the Leibniz formula for the determinant of a 3×3 matrix. Recall that

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} \\ - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \\ = \sum_{\sigma \in S_3} \operatorname{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} a_{3\sigma(3)} ,$$

where

$$\sigma = (\sigma(1), \sigma(2), \sigma(3)) \in S_3 := \{\text{permutations of } 3 \text{ elements}\} \\ = \{(1,2,3),(1,3,2),(2,3,1),(2,1,3),(3,1,2),(3,2,1)\} ,$$

and the corresponding signs flip each time two elements get swapped:

$$\operatorname{sign}(1,2,3) = 1 , \quad \operatorname{sign}(1,3,2) = -1 , \quad \operatorname{sign}(3,1,2) = 1 , \\ \operatorname{sign}(3,2,1) = -1 , \quad \operatorname{sign}(2,3,1) = 1 , \quad \operatorname{sign}(2,1,3) = -1 .$$

An **even permutation** is a permutation σ with $\operatorname{sign}(\sigma) = 1$; an **odd permutation** is a permutation σ with $\operatorname{sign}(\sigma) = -1$.

EXAMPLES 2.18. Let $V = \mathbb{R}[x]_2$.

(1) Let $p, q \in \mathbb{R}[x]_2$. The function $\varphi(p,q) := p(\pi)q(33)$ is a bilinear form.

(2) Likewise,

$$\varphi(p,q) := p'(0)q(4) - 5p'(3)q''(\tfrac{1}{2})$$

is a bilinear form.

□

EXERCISE 2.19. Are the following functions bilinear forms?

(1) $V = \mathbb{R}^2$ and $\varphi(u,v) := \det \begin{bmatrix} u \\ v \end{bmatrix}$;

(2) $V = \mathbb{R}[x]_2$ and $\varphi(p,q) := \int_0^1 p(x)q(x)dx$;

(3) $V = M_{2\times 2}(\mathbb{R})$, the space of real 2×2 matrices, and $\varphi(L,M) := L_1^1 \operatorname{tr} M$, where L_1^1 it the (1,1)-entry of L and $\operatorname{tr} M$ is the trace of M;

(4) $V = \mathbb{R}^3$ and $\varphi(v,w) := v \times w$;

(5) $V = \mathbb{R}^2$ and $\varphi(v,w)$ is the area of the parallelogram spanned by v and w.

(6) $V = M_{n\times n}(\mathbb{R})$, the space of real $n \times n$ matrices with $n > 1$, and $\varphi(L,M) := \operatorname{tr} L \det M$, where $\operatorname{tr} L$ is the trace of L and $\det M$ is the determinant of M.

REMARK 2.20. We need to be careful about the following possible confusion. A bilinear form on V is a function on $V \times V$ that is linear in each variable *separately*. But $V \times V$ is also a vector space and one might wonder whether a bilinear form on V is also a linear form on the vector space $V \times V$. But this is not the case. For example consider the case in which $V = \mathbb{R}$, so that $V \times V = \mathbb{R}^2$ and let $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a function:

(1) If $\varphi(x, y) := 2x - y$, then φ is not a *bilinear form on* $\mathbb{R}$, but is a *linear form* on $(x, y) \in \mathbb{R}^2$;
(2) If $\varphi(x, y) := 2xy$, then φ is a *bilinear form on* $\mathbb{R}$ (hence *linear* in $x \in \mathbb{R}$ and *linear* in $y \in \mathbb{R}$), but it is *not* a linear form on $\mathbb{R}^2$, as it is *not linear* in $(x, y) \in \mathbb{R}^2$.

So a *bi*linear form is not a form that it is "twice as linear" as a linear form, but a form that is defined on the product of twice the vector space.

□

EXERCISE 2.21. Verify the above assertions in Remark 2.20 to make sure you understand the difference.

2.2.2. Tensor Product of Two Linear Forms on V.

Let $\alpha, \beta \in V^*$ be two linear forms, $\alpha, \beta : V \to \mathbb{R}$, and define $\varphi : V \times V \to \mathbb{R}$, by

$$\varphi(v, w) := \alpha(v)\beta(w) .$$

Then φ is bilinear, is called the **tensor product** of α and β and is denoted by

$$\varphi = \alpha \otimes \beta .$$

NOTE 2.22. In general $\alpha \otimes \beta \neq \beta \otimes \alpha$, as there could be vectors v and w such that $\alpha(v)\beta(w) \neq \beta(v)\alpha(w)$.

EXAMPLE 2.23. Let $V = \mathbb{R}[x]_2$, let $\alpha(p) = p(2) - p'(2)$ and $\beta(p) = \int_3^4 p(x)dx$ be two linear forms. Then

$$(\alpha \otimes \beta)(p, q) = (p(2) - p'(2)) \int_3^4 q(x)dx$$

is a bilinear form. □

REMARK 2.24. The bilinear form $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by the formula $\varphi(x, y) := 2xy$ is the tensor product of two linear forms on $\mathbb{R}$, for instance, $\varphi(x, y) = (\alpha \otimes \alpha)(x, y)$ where $\alpha : \mathbb{R} \to \mathbb{R}$ is the linear form given by $\alpha(x) := \sqrt{2}x$.

On the other hand, not every bilinear form is simply the tensor product of two linear forms. As we will see below, the first such examples are found for bilinear forms on vector spaces of dimension at least 2. □

2.2.3. A Basis for Bilinear Forms.

Let

$$\mathrm{Bil}(V \times V, \mathbb{R}) := \{\text{all bilinear forms } \varphi : V \times V \to \mathbb{R}\} .$$

EXERCISE 2.25. Check that $\mathrm{Bil}(V\times V,\mathbb{R})$ is a vector space with the zero element equal to the bilinear form identically equal to zero.
Hint: It is enough to check that if $\varphi,\psi\in\mathrm{Bil}(V\times V,\mathbb{R})$, and $\lambda,\mu\in\mathbb{R}$, then $\lambda\varphi+\mu\psi\in\mathrm{Bil}(V\times V,\mathbb{R})$. Why? (Recall Example 1.3(3) on page 8 and Exercise 2.7 on page 29.)

Assuming Exercise 2.25, we are going to find a basis of $\mathrm{Bil}(V\times V,\mathbb{R})$ and determine its dimension. Let $\mathcal{B}=\{b_1,\dots,b_n\}$ be a basis of V and let $\mathcal{B}^*=\{\beta^1,\dots,\beta^n\}$ be the dual basis of V^* (that is $\beta^i(b_j)=\delta^i_j$).

PROPOSITION 2.26. *The bilinear forms* $\beta^i\otimes\beta^j$*,* $i,j=1,\dots,n$ *form a basis of* $\mathrm{Bil}(V\times V,\mathbb{R})$*. As a consequence,* $\dim\mathrm{Bil}(V\times V,\mathbb{R})=n^2$.

NOTATION. We denote

$$\boxed{\mathrm{Bil}(V\times V,\mathbb{R})=V^*\otimes V^*}$$

and call this vector space the **tensor product** of V^* and V^*. A justification for this notation will appear in §4.3.2.

REMARK 2.27. Just as it is for linear forms, to verify that two bilinear forms on V are the same it is enough to verify that they are the same on every pair of elements of a basis of V. In fact, let φ,ψ be two bilinear forms, let $\mathcal{B}=\{b_1,\dots,b_n\}$ be a basis of V, and assume that

$$\varphi(b_i,b_j)=\psi(b_i,b_j)$$

for all $1\le i,j,\le n$. Let $v=v^ib_i, w=w^jb_j\in V$ be arbitrary vectors. We now verify that $\varphi(v,w)=\psi(v,w)$. Because of the linearity in each variable, we have

$$\varphi(v,w)=\varphi(v^ib_i,w^jb_j)=v^iw^j\varphi(b_i,b_j)=v^iw^j\psi(b_i,b_j)=\psi(v^ib_i,w^jb_j)=\psi(v,w)\,.$$

□

PROOF OF PROPOSITION 2.26. The proof will be similar to the one of Proposition 2.9 for linear forms. We first check that the set of bilinear forms $\{\beta^i\otimes\beta^j,\, i,j=1,\dots,n\}$ consists of linearly independent vectors, then that it spans $\mathrm{Bil}(V\times V,\mathbb{R})$.

For the linear independence we need to check that the only linear combination of the $\beta^i\otimes\beta^j$ that gives the zero bilinear form is the trivial linear combination. Let $c_{ij}\beta^i\otimes\beta^j=0$ be a linear combination of the $\beta^i\otimes\beta^j$. Then for all pairs of basis vectors (b_k,b_ℓ), with $k,\ell=1,\dots,n$, we have

$$0=c_{ij}\beta^i\otimes\beta^j(b_k,b_\ell)=c_{ij}\delta^i_k\delta^j_\ell=c_{k\ell}\,,$$

thus showing the linear independence.

To check that $\mathrm{span}\{\beta^i\otimes\beta^j,\, i,j=1,\dots,n\}=\mathrm{Bil}(V\times V,\mathbb{R})$, we need to check that if $\varphi\in\mathrm{Bil}(V\times V,\mathbb{R})$, there exists $B_{ij}\in\mathbb{R}$ such that

$$\varphi=B_{ij}\beta^i\otimes\beta^j\,.$$

Because of (2.2) on page 28, we obtain

$$\varphi(b_k,b_\ell)=B_{ij}\beta^i(b_k)\beta^j(b_\ell)=B_{ij}\delta^i_k\delta^j_\ell=B_{k\ell}\,,$$

for every pair $(b_k, b_\ell) \in V \times V$. Hence, we set $B_{k\ell} := \varphi(b_k, b_\ell)$. Now both φ and $\varphi(b_k, b_\ell)\beta^i \otimes \beta^j$ are bilinear forms and they coincide on $\mathcal{B} \times \mathcal{B}$. Because of the above Remark 2.27, the two bilinear forms coincide. □

EXAMPLE 2.28. We continue with the study of the *scalar triple product* $\varphi_u : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$, that was defined in Example 2.16 for a fixed given vector $u = \begin{bmatrix} u^1 \\ u^2 \\ u^3 \end{bmatrix}$. We now want to find the components B_{ij} of φ_u with respect to the standard basis of $\mathbb{R}^3$.

Recall the cross product in $\mathbb{R}^3$ is defined on the elements of the standard basis by

$$e_i \times e_j := \begin{cases} 0 & \text{if } i = j \\ e_k & \text{if } (i,j,k) \text{ is a cyclic permutation of } (1,2,3) \\ -e_k & \text{if } (i,j,k) \text{ is a noncyclic permutation of } (1,2,3)\,, \end{cases}$$

that is

$$\text{cyclic}\begin{cases} e_1 \times e_2 = e_3 \\ e_2 \times e_3 = e_1 \\ e_3 \times e_1 = e_2 \end{cases} \quad \text{and} \quad \text{noncyclic}\begin{cases} e_2 \times e_1 = -e_3 \\ e_3 \times e_2 = -e_1 \\ e_1 \times e_3 = -e_2 \end{cases}$$

Since $u \cdot e_k = u^k$, then

$$B_{ij} = \varphi_u(e_i, e_j) = u \cdot (e_i \times e_j) = \begin{cases} 0 & \text{if } i = j \\ u^k & \text{if } (i,j,k) \text{ is a cyclic permutation of } (1,2,3) \\ -u^k & \text{if } (i,j,k) \text{ is a noncyclic permutation of } (1,2,3) \end{cases}$$

Thus

$$\begin{aligned} &B_{12} = u^3 = -B_{21} \\ &B_{31} = u^2 = -B_{13} \\ &B_{23} = u^1 = -B_{32} \\ &B_{11} = B_{22} = B_{33} = 0 \quad \text{(that is, the diagonal components are zero)}\,, \end{aligned}$$

which can be written as a matrix

$$B = \begin{bmatrix} 0 & u^3 & -u^2 \\ -u^3 & 0 & u^1 \\ u^2 & -u^1 & 0 \end{bmatrix}.$$

The components B_{ij} of B are the components of this bilinear form with respect to the basis $\beta^i \otimes \beta^j$ $(i, j = 1, \ldots, n)$, where $\beta^i(e_k) = \delta^i_k$. Hence, we can write

$$\begin{aligned} \varphi_u = B_{ij}\beta^i \otimes \beta^j \;&=\; u_1\left(\beta^2 \otimes \beta^3 - \beta^3 \otimes \beta^2\right) \\ &\quad + u_2\left(\beta^3 \otimes \beta^1 - \beta^1 \otimes \beta^3\right) + u_3\left(\beta^1 \otimes \beta^2 - \beta^2 \otimes \beta^1\right)\,. \end{aligned}$$

□

2.2.4. Covariance of Bilinear Forms.

We have seen that, once we choose a basis $\mathcal{B} = \{b_1, \dots, b_n\}$ of V, we automatically have a basis $\mathcal{B}^* = \{\beta^1, \dots, \beta^n\}$ of V^* and a basis $\{\beta^i \otimes \beta^j,\ i, j = 1, \dots, n\}$ of $V^* \otimes V^*$. This implies, that any bilinear form $\varphi : V \times V \to \mathbb{R}$ can be represented by its components

$$\boxed{B_{ij} = \varphi(b_i, b_j)}, \tag{2.13}$$

in the sense that

$$\boxed{\varphi = B_{ij}\beta^i \otimes \beta^j}.$$

Moreover, these components can be arranged in a matrix[6]

$$B := \begin{bmatrix} B_{11} & \dots & B_{1n} \\ \vdots & & \vdots \\ B_{n1} & \dots & B_{nn} \end{bmatrix}$$

called the **matrix of the bilinear form** φ with respect to the chosen basis $\mathcal{B}$. The natural question of course is: how does the matrix B change when we choose a different basis of V?

So, let us choose a different basis $\widetilde{\mathcal{B}} := \{\tilde{b}_1, \dots, \tilde{b}_n\}$ and corresponding bases $\widetilde{\mathcal{B}}^* = \{\tilde{\beta}^1, \dots, \tilde{\beta}^n\}$ of V^* and $\{\tilde{\beta}^i \otimes \tilde{\beta}^j,\ i, j = 1, \dots, n\}$ of $V^* \otimes V^*$, with respect to which φ will be represented by a matrix $\widetilde{B}$, whose entries are $\widetilde{B}_{ij} = \varphi(\tilde{b}_i, \tilde{b}_j)$.

To see the relation between B and $\widetilde{B}$, due to the change of basis from $\mathcal{B}$ to $\widetilde{\mathcal{B}}$, we start with the matrix of the change of basis $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$, according to which

$$\tilde{b}_j = L^i_j b_i\,. \tag{2.14}$$

Then

$$\widetilde{B}_{ij} = \varphi(\tilde{b}_i, \tilde{b}_j) = \varphi(L^k_i b_k, L^\ell_j b_\ell) = L^k_i L^\ell_j \varphi(b_k, b_\ell) = L^k_i L^\ell_j B_{k\ell}\,,$$

where the first and the last equality follow from (2.13), the second from (2.14) (after having renamed the dummy indices to avoid conflicts) and the remaining one from the bilinearity of σ. We conclude that

$$\boxed{\widetilde{B}_{ij} = L^k_i L^\ell_j B_{k\ell}}.$$

EXERCISE 2.29. Show that the formula of the transformation of the component of a bilinear form in terms of the matrices of the change of coordinates is

$$\boxed{\widetilde{B} = {}^{\mathsf{t}}LBL}, \tag{2.15}$$

where ${}^{\mathsf{t}}L$ denotes the transpose of the matrix L.

[6]Note that, contrary to the matrix that gives the change of coordinates between two basis of the vector space, here we have only lower indices. This is not by chance and reflects the type of tensor a bilinear form is.

We hence say that a bilinear form φ is a **covariant 2-tensor** or a **tensor of type** $(0, 2)$.

2.3. Multilinear Forms

2.3.1. Definition, Basis and Covariance.

We saw in §2.1.3 that linear forms are covariant 1-tensors – or tensors of type $(0, 1)$ – and in §2.2.4 that bilinear forms are covariant 2-tensors – or tensors of type $(0, 2)$.

Analogously to what was done until now, one can define **trilinear forms** on V, that is functions $T : V \times V \times V \to \mathbb{R}$ that are linear with respect to each of the three arguments. The space of trilinear forms on V is denoted

$$V^* \otimes V^* \otimes V^* ,$$

has basis

$$\{\beta^j \otimes \beta^j \otimes \beta^k,\, i, j, k = 1, \ldots, n\}$$

and, hence, has dimension n^3. The tensor product $\otimes$ is defined as above.

Since the components of a trilinear form $T : V \times V \times V \to \mathbb{R}$ satisfy the following transformation with respect to a change of basis

$$\widetilde{T}_{ijk} = L_i^\ell L_j^p L_k^q T_{\ell pq} ,$$

a trilinear form is a **covariant** 3**-tensor** or a **tensor of type** $(0, 3)$**.**

Of course, there is nothing special about $k = 1, 2$ or 3:

Definition 2.30. A k-**linear form** or **multilinear form of order** k on V is a function $f : V \times \cdots \times V \to \mathbb{R}$ from k-copies of V into $\mathbb{R}$, that is linear in each of its arguments.

A k-linear form is a **covariant** k**-tensor** (or a **covariant tensor of order** k or a **tensor of type** $(0, k)$). The vectors space of k-linear forms on V, denoted

$$\underbrace{V^* \otimes \cdots \otimes V^*}_{k \text{ factors}},$$

has basis

$$\beta^{i_1} \otimes \beta^{i_2} \otimes \cdots \otimes \beta^{i_k} , \quad i_1, \ldots, i_k = 1, \ldots, n$$

and, hence, $\dim(V^* \otimes \cdots \otimes V^*) = n^k$.

2.3.2. Examples of Multilinear Forms.

Example 2.31. We once more address the *scalar triple product*,[7] discussed in Examples 2.16 and 2.28. This time we want to find the components B_{ij} of φ_u with respect

[7]The scalar triple product is called *Spatprodukt* in German.

to the (nonstandard) basis

$$\widetilde{\mathcal{B}} := \left\{ \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\tilde{b}_1}, \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\tilde{b}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\tilde{b}_3} \right\}.$$

The matrix of the change of coordinates from the standard basis to $\widetilde{\mathcal{B}}$ is

$$L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix},$$

so that

$$\begin{aligned} \widetilde{B} &= \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}}_{{}^{t}L} \underbrace{\begin{bmatrix} 0 & u^3 & -u^2 \\ -u^3 & 0 & u^1 \\ u^2 & -u^1 & 0 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}}_{L} \\ &= \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}}_{{}^{t}L} \underbrace{\begin{bmatrix} u^3 & -u^2 & -u^2 \\ 0 & u^1 - u^3 & u^1 \\ -u^1 & u^2 & 0 \end{bmatrix}}_{BL} = \begin{bmatrix} 0 & u^1 - u^3 & u^1 \\ u^3 - u^1 & 0 & -u^2 \\ -u^1 & u^2 & 0 \end{bmatrix}. \end{aligned}$$

It is easy to check that $\widetilde{B}$ is antisymmetric just like B is, and to check that the components of $\widetilde{B}$ are correct by using the formula for φ. In fact

$$\begin{aligned} \widetilde{B}_{12} &= \varphi(\tilde{b}_1, \tilde{b}_2) = u \cdot (e_2 \times (e_1 + e_3)) = u^1 - u^3 \\ \widetilde{B}_{13} &= \varphi(\tilde{b}_1, \tilde{b}_3) = u \cdot ((e_2) \times e_3) = u^1 \\ \widetilde{B}_{23} &= \varphi(\tilde{b}_2, \tilde{b}_3) = u \cdot ((e_1 + e_3) \times e_3) = -u^2 \\ \widetilde{B}_{11} &= \varphi(\tilde{b}_1, b_1) = u \cdot (e_2 \times e_2) = 0 \\ \widetilde{B}_{22} &= \varphi(\tilde{b}_2, b_2) = u \cdot ((e_1 + e_3) \times (e_1 + e_3)) = 0 \\ \widetilde{B}_{33} &= \varphi(\tilde{b}_3, b_3) = u \cdot (e_3 \times e_3) = 0 \end{aligned}$$

□

EXAMPLE 2.32. If, in the definition of the scalar triple product, instead of fixing a vector $a \in \mathbb{R}$, we let the vector vary, we have a function $\varphi : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$, defined by

$$\varphi(u, v, w) := u \cdot (v \times w) = \det \begin{bmatrix} u \\ v \\ w \end{bmatrix}.$$

One can verify that such function is trilinear, that is linear in each of the three variables separately.

The components T_{ijk} of this trilinear form are simply given by the sign of the corresponding permutation:

$$\begin{aligned}\varphi = \operatorname{sign}(i,j,k)\beta^i \otimes \beta^j \otimes \beta^k \ &= \ \beta^1 \otimes \beta^2 \otimes \beta^3 - \beta^1 \otimes \beta^3 \otimes \beta^2 + \beta^3 \otimes \beta^1 \otimes \beta^2 \\ &\quad -\beta^3 \otimes \beta^2 \otimes \beta^1 + \beta^2 \otimes \beta^3 \otimes \beta^1 - \beta^2 \otimes \beta^1 \otimes \beta^3 \,,\end{aligned}$$

where the sign of the permutation is given by

$$\operatorname{sign}(i,j,k) \ := \ \begin{cases} +1 & \text{if } (i,j,k) = (1,2,3), (2,3,1) \text{ or } (3,1,2) \\ & \qquad\qquad\quad (\textbf{even} \text{ permutations of } (1,2,3)) \\ -1 & \text{if } (i,j,k) = (1,3,2), (2,1,3) \text{ or } (3,2,1) \\ & \qquad\qquad\quad (\textbf{odd} \text{ permutations of } (1,2,3)) \\ 0 & \text{otherwise.} \end{cases}$$

□

EXAMPLE 2.33. In general, the **determinant** defines an n-linear form in $\mathbb{R}^n$ by

$$\varphi : \underbrace{\mathbb{R}^n \times \ldots \times \mathbb{R}^n}_{n \text{ factors}} \longrightarrow \mathbb{R}\,, \qquad \varphi(v_1, \ldots, v_n) := \det \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}\,,$$

where we compute the determinant of the square matrix with rows (equivalently, columns) given by the n vectors. Multilinearity is a fundamental property of the determinant.

In this case, the components of this multilinear form are also given by the permutation signs:

$$\varphi = \operatorname{sign}(i_1, \ldots, i_n)\beta^{i_1} \otimes \ldots \otimes \beta^{i_n}\,,$$

where

$$\operatorname{sign}(i_1, \ldots, i_n) \ := \ \begin{cases} +1 & \text{if } (i_1, \ldots, i_n) \text{ is an even permutation of } (1, \ldots, n) \\ -1 & \text{if } (i_1, \ldots, i_n) \text{ is an odd permutation of } (1, \ldots, n) \\ 0 & \text{otherwise.} \end{cases}$$

A permutation of $(1, 2, \ldots, n)$ is called an **even permutation**, if it is obtained from $(1, 2, \ldots, n)$ by an even number of two-element swaps; otherwise it is called an **odd permutation**. □

2.3.3. Tensor Product for Multilinear Forms.

Let

$$T : \underbrace{V \times \cdots \times V}_{k \text{ times}} \to \mathbb{R} \qquad \text{and} \qquad U : \underbrace{V \times \cdots \times V}_{\ell \text{ times}} \to \mathbb{R}$$

be, respectively, a k-linear and an ℓ-linear form. Then the **tensor product** of T and U is the function

$$T \otimes U : \underbrace{V \times \cdots \times V}_{k+\ell \text{ times}} \to \mathbb{R}$$

defined by

$$T \otimes U(v_1, \ldots, v_{k+\ell}) := T(v_1, \ldots, v_k)U(v_{k+1}, \ldots, v_{k+\ell}).$$

This is a $(k+\ell)$-linear form. Equivalently, this is saying that the tensor product of a tensor of type $(0, k)$ and a tensor of type $(0, \ell)$ is a tensor of type $(0, k+\ell)$. Later we will see how this product extends to more general tensors.

CHAPTER 3

Inner Products

3.1. Definitions and First Properties

Inner products are a special case of bilinear forms. They add an important structure to a vector space, as for example they allow to compute the length of a vector and they provide a canonical identification between the vector space V and its dual V^*.

3.1.1. Inner Products and their Related Notions.

DEFINITION 3.1. An **inner product** $g : V \times V \to \mathbb{R}$ on a real vector space V is a *bilinear form* on V that is

(1) *symmetric*, that is $g(v, w) = g(w, v)$ for all $v, w \in V$ and
(2) *positive definite*, that is $g(v, v) \geq 0$ for all $v \in V$, and $g(v) = 0$ if and only if $v = 0$.

EXERCISE 3.2. Let $V = \mathbb{R}^3$. Verify that the dot product $\varphi(v, w) := v \cdot w$, defined as

$$v \cdot w = v^i w^i \,,$$

where $v = \begin{bmatrix} v^1 \\ v^2 \\ v^3 \end{bmatrix}$ and $w = \begin{bmatrix} w^1 \\ w^2 \\ w^3 \end{bmatrix}$ is an inner product. This is called the **standard inner product**.

EXERCISE 3.3. Determine whether the following bilinear forms $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ are inner products, by verifying whether they are symmetric and positive definite (the formulas are troughout defined for all $v, w \in \mathbb{R}^n$):

(1) $\varphi(v, w) := -v \cdot w;$
(2) $\varphi(v, w) := v \cdot w + 2v^1 w^2;$
(3) $\varphi(v, w) := v^1 w^1;$
(4) $\varphi(v, w) := v \cdot w - 2v^1 w^1;$
(5) $\varphi(v, w) := v \cdot w + 2v^1 w^1;$
(6) $\varphi(v, w) := v \cdot 3w.$

EXERCISE 3.4. Let $V := \mathbb{R}[x]_2$ be the vector space of polynomials of degree ≤ 2. Determine whether the following bilinear forms are inner products, by verifying whether they are symmetric and positive definite:

(1) $\varphi(p,q) := \int_0^1 p(x)q(x)dx$;
(2) $\varphi(p,q) := \int_0^1 p'(x)q'(x)dx$;
(3) $\varphi(p,q) := \int_3^\pi e^x p(x)q(x)dx$;
(4) $\varphi(p,q) := p(1)q(1) + p(2)q(2)$;
(5) $\varphi(p,q) := p(1)q(1) + p(2)q(2) + p(3)q(3)$.
(6) $\varphi(p,q) := p(1)q(2) + p(2)q(3) + p(3)q(1)$.

Definition 3.5. Let $g : V \times V \to \mathbb{R}$ be an inner product on V.

(1) The **norm** $\|v\|$ of a vector $v \in V$ is defined as
$$\|v\| := \sqrt{g(v,v)}\,.$$
(2) A vector $v \in V$ is **unit vector** if $\|v\| = 1$;
(3) Two vectors $v, w \in V$ are **orthogonal** (that is, **perpendicular** denoted $v \perp w$), if $g(v,w) = 0$;
(4) Two vectors $v, w \in V$ are **orthonormal** if they are orthogonal and $\|v\| = \|w\| = 1$;
(5) A basis $\mathcal{B}$ of V is an **orthonormal basis** if $b_1, \dots, b_n$ are pairwise orthonormal vectors, that is
$$g(b_i, b_j) = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j\,, \end{cases} \tag{3.1}$$
for all $i, j = 1 \dots, n$. The condition for $i = j$ implies that an orthonormal basis consists of unit vectors, while the one for $i \neq j$ implies that it consists of pairwise orthogonal vectors.

Example 3.6.

(1) Let $V = \mathbb{R}^n$ and g the standard inner product. The standard basis $\mathcal{B} = \{e_1, \dots, e_n\}$ is an orthonormal basis with respect to the standard inner product.
(2) Let $V = \mathbb{R}[x]_2$ and let $g(p,q) := \int_{-1}^1 p(x)q(x)dx$. Check that the basis
$$\mathcal{B} = \{p_1, p_2, p_3\}\,,$$
where
$$p_1(x) := \frac{1}{\sqrt{2}}, \quad p_2(x) := \sqrt{\frac{3}{2}}x, \quad p_3(x) := \sqrt{\frac{5}{8}}(3x^2 - 1)\,,$$
is an orthonormal basis with respect to the inner product g. Up to scaling, p_1, p_2, p_3 are the so-called first three **Legendre polynomials**.

□

An inner product g on a vector space V induces a **metric**[8] on V, where the distance between vectors $v, w \in V$ is given by

$$d(v,w) := \|v - w\| .$$

3.1.2. Symmetric Matrices and Quadratic Forms.

Recall that a matrix $S \in M_{n\times n}(\mathbb{R})$ is **symmetric** if $S = {}^{\mathsf{t}}S$, that is if

$$S = \begin{bmatrix} * & a & b & \cdots \\ a & * & c & \cdots \\ b & c & * & \cdots \\ \cdots & \cdots & & * \end{bmatrix} .$$

Moreover, if S is symmetric, then

(1) S is **positive definite** if ${}^{\mathsf{t}}vSv > 0$ for all $v \in \mathbb{R}^n \setminus \{0\}$;
(2) S is **negative definite** if ${}^{\mathsf{t}}vSv < 0$ for all $v \in \mathbb{R}^n \setminus \{0\}$;
(3) S is **positive semidefinite** if ${}^{\mathsf{t}}vSv \geq 0$ for all $v \in \mathbb{R}^n$;
(4) S is **positive semidefinite** if ${}^{\mathsf{t}}vSv \leq 0$ for all $v \in \mathbb{R}^n$;
(5) S is **indefinite** if $v^{\mathsf{t}}Sv$ takes both positive and negative values for different $v \in \mathbb{R}^n$.

Definition 3.7. A **quadratic form** $Q : \mathbb{R}^n \to \mathbb{R}$ is a homogeneous quadratic polynomial in n variables:

$$Q(x^1, \dots, x^n) = Q_{ij}x^i x^j , \qquad \text{where } Q_{ij} \in \mathbb{R} .$$

To any symmetric matrix S corresponds a **quadratic form** $Q_S : \mathbb{R}^n \to \mathbb{R}$ defined by

$$Q_S(v) = {}^{\mathsf{t}}vSv = \underbrace{\begin{bmatrix} v^1 & \dots & v^n \end{bmatrix} S \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix}}_{\text{matrix notation}} = \underbrace{v^i v^j S_{ij}}_{\text{Einstein notation}} . \tag{3.2}$$

Note that Q is *not* linear in v.

Let S be a symmetric matrix and Q_S be the corresponding quadratic form. The notion of positive definiteness, etc. for S can be translated into corresponding properties for Q_S, namely:

(1) Q is **positive definite** if $Q(v) > 0$ for all $v \in \mathbb{R}^n \setminus \{0\}$;
(2) Q is **negative definite** if $Q(v) < 0$ for all $v \in \mathbb{R}^n \setminus \{0\}$;

[8]An inner product induces a *norm* and a norm induces a *metric* on a vector space. However, the converses do not hold.

(3) Q is **positive semidefinite** if $Q(v) \geq 0$ for all $v \in \mathbb{R}^n$;
(4) Q is **negative semidefinite** if $Q(v) \leq 0$ for all $v \in \mathbb{R}^n$;
(5) Q is **indefinite** if $Q(v)$ takes both positive and negative values.

EXAMPLE 3.8. We consider $\mathbb{R}^2$ with the standard basis $\mathcal{E}$ and the quadratic form[9] $Q(v) := v^1v^1 - v^2v^2$, where $v = \begin{bmatrix} v^1 \\ v^2 \end{bmatrix}$. The symmetric matrix corresponding to Q is $S := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. If $v = \begin{bmatrix} v^1 \\ 0 \end{bmatrix}$, then $Q(v) = 1 > 0$, if $v = \begin{bmatrix} 0 \\ v^2 \end{bmatrix}$, then $Q(v) = -1 < 0$, but any vector for which $v^1 = v^2$ has the property that $Q(v) = 0$. □

To find out the type of a symmetric matrix S (or, equivalently of a quadratic form Q_S) it is enough to look at the eigenvalues of S, namely:

(1) S and Q_S are *positive definite* when all eigenvalues of S are positive:
(2) S and Q_S are *negative definite* when all eigenvalues of S are negative;
(3) S and Q_S are *positive semidefinite* when all eigenvalues of S are non-negative;
(4) S and Q_S are *negative semidefinite* when all eigenvalues of S are non-positive;
(5) S and Q_S are *indefinite* when S has both positive and negative eigenvalues.

The reason this makes sense is the same reason for which we need to restrict our attention to symmetric matrices and lies in the so-called Spectral Theorem:

THEOREM 3.9. *[Spectral Theorem] Any $n \times n$ symmetric matrix S has the following properties:*

(1) *it has only* real *eigenvalues;*
(2) *it is* diagonalizable*;*
(3) *it admits an* orthonormal eigenbasis*, that is, a basis $\{b_1, \dots, b_n\}$ of $\mathbb{R}^n$ such that the b_j are orthonormal and are eigenvectors of S.*

3.1.3. Inner Products vs. Symmetric Positive Definite Matrices.

Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis of V and g an inner product. The **components of g with respect to $\mathcal{B}$** are

$$g_{ij} := g(b_i, b_j)\,. \tag{3.3}$$

Let G be the matrix with entries g_{ij}

$$G = \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nn} \end{bmatrix}\,. \tag{3.4}$$

We claim that G is symmetric and positive definite. In fact:

[9]Here, we avoid the usual notation for squares, because of the possible confusion with upper indices.

(1) Since g is *symmetric*, then for $1 \leq i, j \leq n$,

$$g_{ij} = g(b_i, b_j) = g(b_j, b_i) = g_{ji} \quad \Longrightarrow \quad G \text{ is a } \textit{symmetric} \text{ matrix;} \tag{3.5}$$

(2) Since g is *positive definite*, then G is *positive definite* as a symmetric matrix. In fact, let $v = v^i b_i, w = w^j b_j \in V$ be two vectors. Then, using the bilinearity of g, the definition (3.3) and Einstein notation, we have:

$$g(v, w) = g(v^i b_i, w^j b_j) = v^i w^j \underbrace{g(b_i, b_j)}_{g_{ij}} = v^i w^j g_{ij}$$

or, in matrix notation,

$$g(v, w) = {}^{\mathsf{t}}[v]_{\mathcal{B}} G [w]_{\mathcal{B}} = \begin{bmatrix} v^1 & \dots & v^n \end{bmatrix} G \begin{bmatrix} w^1 \\ \vdots \\ w^n \end{bmatrix} .$$

Conversely, if S is a symmetric positive definite matrix and $\mathcal{B} = \{b_1, \dots, b_n\}$ is a basis of V, then the assignment

$$V \times V \longrightarrow \mathbb{R}, \qquad (v, w) \longmapsto {}^{\mathsf{t}}[v]_{\mathcal{B}} S [w]_{\mathcal{B}}$$

defines a map that is seen to be bilinear, symmetric and positive definite, hence an inner product.

3.1.4. Orthonormal Bases.

Suppose that there is a basis $\mathcal{B} = \{b_1, \dots, b_n\}$ of V consisting of orthonormal vectors with respect to g, so that

$$g_{ij} = \delta_{ij} ;$$

cf. Definitions 3.5(5) and (3.3). In other words the symmetric matrix corresponding to the inner product g in the basis consisting of orthonormal vectors is the identity matrix. Moreover

$$g(v, w) = v^i w^j g_{ij} = v^i w^j \delta_{ij} = v^i w^i ,$$

so that, if $v = w$,

$$\|v\|^2 = g(v, v) = v^i v^i = v^1 v^1 + \dots + v^n v^n .$$

We thus deduce that:

FACT 3.10. *Any inner product g can be expressed in the standard form*

$$g(v, w) = v^i w^i ,$$

as long as $[v]_{\mathcal{B}} = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$ *and* $[w]_{\mathcal{B}} = \begin{pmatrix} w^1 \\ \vdots \\ w^n \end{pmatrix}$ *are the coordinates of v and w with respect to an orthonormal basis $\mathcal{B}$ for g.*

EXAMPLE 3.11. Let g be an inner product of $\mathbb{R}^3$ with respect to which

$$\widetilde{\mathcal{B}} := \left\{ \underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix}}_{\tilde{b}_1}, \underbrace{\begin{bmatrix}1\\1\\0\end{bmatrix}}_{\tilde{b}_2}, \underbrace{\begin{bmatrix}1\\1\\1\end{bmatrix}}_{\tilde{b}_3} \right\}$$

is an orthonormal basis. We want to express g with respect to the standard basis $\mathcal{E}$ of $\mathbb{R}^3$

$$\mathcal{E} := \left\{ \underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix}}_{e_1}, \underbrace{\begin{bmatrix}0\\1\\0\end{bmatrix}}_{e_2}, \underbrace{\begin{bmatrix}0\\0\\1\end{bmatrix}}_{e_3} \right\}.$$

The matrices of the change of basis are

$$L := L_{\widetilde{\mathcal{B}}\mathcal{E}} = \begin{bmatrix}1&1&1\\0&1&1\\0&0&1\end{bmatrix} \quad \text{and } \Lambda = L^{-1} = \begin{bmatrix}1&-1&0\\0&1&-1\\0&0&1\end{bmatrix}.$$

Since g is a bilinear form, we saw in (2.15) that its matrices with respect to the bases $\mathcal{B}$ and $\mathcal{E}$ are related by the formula

$$\widetilde{G} = {}^{\mathsf{t}}LGL\,.$$

Since the basis $\widetilde{\mathcal{B}}$ is orthonormal with respect to g, the associated matrix $\widetilde{G}$ is the identity matrix, so that

$$\begin{aligned} G &= {}^{\mathsf{t}}\Lambda\widetilde{G}\Lambda = {}^{\mathsf{t}}\Lambda\Lambda \\ &= \begin{bmatrix}1&0&0\\-1&1&0\\0&-1&1\end{bmatrix}\begin{bmatrix}1&-1&0\\0&1&-1\\0&0&1\end{bmatrix} = \begin{bmatrix}1&-1&0\\-1&2&-1\\0&-1&2\end{bmatrix}. \end{aligned} \tag{3.6}$$

It follows that, with respect to the standard basis, g is given by

$$\begin{aligned} g(v,w) &= \begin{pmatrix}v^1 & v^2 & v^3\end{pmatrix}\begin{bmatrix}1&-1&0\\-1&2&-1\\0&-1&2\end{bmatrix}\begin{pmatrix}w^1\\w^2\\w^3\end{pmatrix} \\ &= v^1w^1 - v^1w^2 - v^2w^1 + 2v^2w^2 \\ &\qquad -v^2w^3 - v^3w^2 + 2v^3w^3\,. \end{aligned} \tag{3.7}$$

□

EXERCISE 3.12. Verify the formula (3.7) for the inner product g in the coordinates of the basis $\widetilde{\mathcal{B}}$ by applying the matrix of the change of coordinate directly on the coordinates vectors $[v]_{\mathcal{E}}$.

REMARK 3.13. The norm and the value of the inner product of vectors depend *only* on the choice of g, but *not* on the choice of basis: different coordinate expressions yield the same result:

$$g(v,w) = {}^{\mathtt{t}}[v]_{\mathcal{B}} G [v]_{\mathcal{B}} = {}^{\mathtt{t}}[v]_{\widetilde{\mathcal{B}}} \widetilde{G} [v]_{\widetilde{\mathcal{B}}}$$

EXAMPLE 3.14. We verify the assertion of the previous remark with the inner product in Example 3.11. Let $v, w \in \mathbb{R}^3$ such that

$$[v]_{\mathcal{E}} = \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} \quad \text{and} \quad [v]_{\widetilde{\mathcal{B}}} = \begin{pmatrix} \tilde{v}^1 \\ \tilde{v}^2 \\ \tilde{v}^3 \end{pmatrix} = L^{-1} \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

and

$$[w]_{\mathcal{E}} = \begin{pmatrix} w^1 \\ w^2 \\ w^3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad \text{and} \quad [w]_{\widetilde{\mathcal{B}}} = \begin{pmatrix} \tilde{w}^1 \\ \tilde{w}^2 \\ \tilde{w}^3 \end{pmatrix} = L^{-1} \begin{pmatrix} w^1 \\ w^2 \\ w^3 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 3 \end{pmatrix} .$$

Then with respect to the basis $\widetilde{\mathcal{B}}$ we have that

$$g(v,w) = 1 \cdot (-1) + 1 \cdot (-1) + 1 \cdot 3 = 1 ,$$

and also with respect to the basis $\mathcal{E}$

$$g(v,w) = 3 \cdot 1 - 3 \cdot 2 - 2 \cdot 1 + 2 \cdot 2 \cdot 2 - 2 \cdot 3 - 1 \cdot 2 + 2 \cdot 1 \cdot 3 = 1 .$$

□

EXERCISE 3.15. Verify that $\|v\| = \sqrt{3}$ and $\|w\| = \sqrt{11}$, when computed with respect to either basis.

Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be an orthonormal basis and let $v = v^i b_i$ be a vector in V. Then the coordinates v^i of v can be obtained from the inner product g and the elements of the basis. In fact, we have

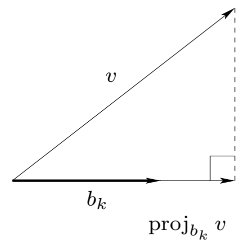
$$g(v, b_j) = g(v^i b_i, b_j) = v^i g(b_i, b_j) = v^i \delta_{ij} = v^j ,$$

that is, the coordinates of a vector with respect to an orthonormal basis are the inner product of the vector with the basis vectors. This is particularly nice, so that we have to make sure that we remember how to construct an orthonormal basis from a given arbitrary basis.

RECALL. The **Gram-Schmidt orthogonalization process** is a recursive process that allows us to obtain an orthonormal basis starting from an arbitrary one. Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be an arbitrary basis, let $g : V \times V \to \mathbb{R}$ be an inner product and $\| \cdot \|$ the corresponding norm.

We start by recalling that the orthogonal projection of a vector $\in V$ onto b_k is defined as

$$\text{(3.8)} \qquad \operatorname{proj}_{b_k} v = \frac{g(v, b_k)}{g(b_k, b_k)} b_k \,.$$



In fact, $\operatorname{proj}_{b_k} v$ is clearly parallel to b_k and the following exercises shows that the component $v - \operatorname{proj}_{b_k} v$ is orthogonal to b_k.

EXERCISE 3.16. With $\operatorname{proj}_{b_k} v$ defined as in (3.8), check that we have

$$v - \operatorname{proj}_{b_k} v \perp b_k \,,$$

where the orthogonality is meant with respect to the inner product g.

Given the basis $\mathcal{B} = \{b_1, \dots, b_n\}$ of V, we will find an orthonormal basis. We start by defining

$$u_1 := \frac{1}{\|b_1\|} b_1 \,.$$

Next, observe that $g(b_2, u_1)u_1$ is the projection of the vector b_2 in the direction of u_1. It follows that

$$b_2^\perp := b_2 - g(b_2, u_1)u_1$$

is a vector orthogonal to u_1, but not necessarily of unit norm. Hence we set

$$u_2 := \frac{1}{\|b_2^\perp\|} b_2^\perp \,.$$

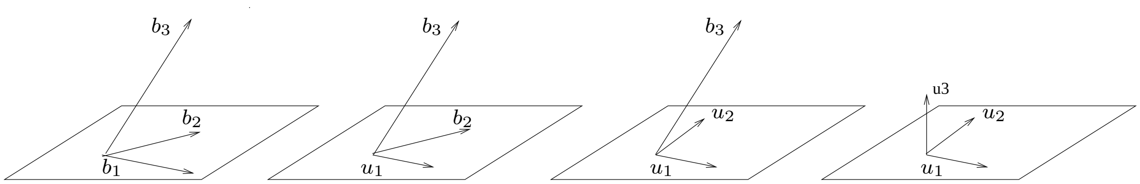
Likewise $g(b_3, u_1)u_1 + (b_3, u_2)u_2$ is the projection of b_3 on the plane generated by u_1 and u_2, so that

$$b_3^\perp := b_3 - g(b_3, u_1)u_1 - g(b_3, u_2)u_2$$

is orthogonal both to u_1 and to u_2. Set

$$u_3 := \frac{1}{\|b_3^\perp\|} b_3^\perp \,.$$

Continuing until we have exhausted all elements of the basis $\mathcal{B}$, we obtain an orthonormal basis $\{u_1, \dots, u_n\}$.



EXAMPLE 3.17. Let V be the subspace of $\mathbb{R}^4$ spanned by

$$b_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \quad b_2 = \begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \end{bmatrix} \quad b_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} .$$

(One can check that b_1, b_2, b_3 are linearly independent and hence form a basis of V.) We look for an orthonormal basis of V with respect to the standard inner product $\langle \cdot , \cdot \rangle$. Since

$$\|b_1\| = (1^1 + 1^2 + (-1)^2 + (-1)^2)^{1/2} = 2 ,$$

we find

$$u_1 := \frac{1}{2} b_1 .$$

Moreover,

$$\langle b_2, u_1 \rangle = \tfrac{1}{2}(1+1) = 2 \qquad \Longrightarrow \qquad b_2^{\perp} := b_2 - \langle b_2, u_1 \rangle u_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} ,$$

so that $\|b_2^{\perp}\| = 2$ and

$$u_2 := \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} .$$

Finally,

$$\langle b_3, u_1 \rangle = \tfrac{1}{2}(1+1-1) = \tfrac{1}{2} \qquad \text{and} \qquad \langle b_3, u_2 \rangle = \tfrac{1}{2}(1+1+1) = \tfrac{3}{2}$$

imply that

$$b_3^{\perp} := b_3 - \langle b_3, u_1 \rangle u_1 - \langle b_3, u_2 \rangle u_2 = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} .$$

Since $\|b_3^{\perp}\| = \frac{\sqrt{2}}{2}$, we have

$$u_3 := \frac{\sqrt{2}}{2} \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} .$$

□

3.2. Reciprocal Basis

3.2.1. Definition and Examples.

Let $g : V \times V \to \mathbb{R}$ be an inner product and $\mathcal{B} = \{b_1, \dots, b_n\}$ any basis of V. From g and $\mathcal{B}$ we can define another basis of V, denoted by

$$\mathcal{B}^g = \{b^1, \dots, b^n\}$$

and satisfying

$$\boxed{g(b^i, b_j) = \delta^i_j}\,. \tag{3.9}$$

The basis $\mathcal{B}^g$ is called the **reciprocal basis** of V with respect to g and $\mathcal{B}$.

Note that, strictly speaking, we are very imprecise here. Indeed, while it is certainly possible to define a set of $n = \dim V$ vectors as in (3.9), we should justify the fact that we call it a *basis*. This will be done in Claim 3.21.

Remark 3.18. In general, $\mathcal{B}^g \neq \mathcal{B}$ and in fact, because of Definition 3.5(5),

$$\mathcal{B} = \mathcal{B}^g \qquad \Longleftrightarrow \qquad \mathcal{B} \text{ is an orthonormal basis.}$$

Example 3.19. Let g be the inner product in (3.7) in Example 3.11 and let $\mathcal{E}$ the standard basis of $\mathbb{R}^3$. We want to find the reciprocal basis $\mathcal{E}^g$, that is we want to find $\mathcal{E}^g := \{e^1, e^2, e^3\}$ such that

$$g(e^i, e_j) = \delta^i_j\,.$$

If G is the matrix of the inner product in (3.6), using the matrix notation and considering e^j as a row vector and e_i as a column vector for $i, j = 1, 2, 3$, we have

$$\begin{bmatrix} — & {}^{t}e^i & — \end{bmatrix} G \begin{bmatrix} | \\ e_j \\ | \end{bmatrix} = \delta^i_j\,.$$

Letting i and j vary from 1 to 3, we obtain

$$\begin{bmatrix} — & {}^{t}e^1 & — \\ — & {}^{t}e^2 & — \\ — & {}^{t}e^3 & — \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} | & | & | \\ e_1 & e_2 & e_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

from which we conclude that

$$\begin{bmatrix} — & {}^{t}e^1 & — \\ — & {}^{t}e^2 & — \\ — & {}^{t}e^3 & — \end{bmatrix} = \begin{bmatrix} | & | & | \\ e_1 & e_2 & e_3 \\ | & | & | \end{bmatrix}^{-1} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} | & | & | \\ e_1 & e_2 & e_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Therefore,

$$e^1 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \quad e^2 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \quad e^3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{3.10}$$

Observe that in order to compute G^{-1} we used the Gauss–Jordan elimination method

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right] \leftrightsquigarrow \left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right]$$

$$\leftrightsquigarrow \left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 2 & 1 & 0 \\ 0 & 1 & -1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]$$

$$\leftrightsquigarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & 2 & 1 \\ 0 & 1 & 0 & 2 & 2 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]$$

□

EXERCISE 3.20. We put ourselves in the situation of Examples 3.19 and 3.11. Namely, let g be an inner product on $\mathbb{R}^3$, let $\mathcal{E} = \{e_1, e_2, e_3\}$ be the standard basis and let

$$\widetilde{\mathcal{B}} := \Bigg\{ \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{\tilde{b}_1}, \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\tilde{b}_2}, \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\tilde{b}_3} \Bigg\}$$

be an orthonormal basis with respect to g.

(1) Compute $[\tilde{b}_1]_{\widetilde{\mathcal{B}}}$, $[\tilde{b}_2]_{\widetilde{\mathcal{B}}}$ and $[\tilde{b}_3]_{\widetilde{\mathcal{B}}}$.
(2) Compute the matrix $G_{\widetilde{\mathcal{B}}}$ associated to g with respect to the basis $\widetilde{\mathcal{B}}$, and the matrix $G_{\mathcal{E}}$ associated to g with respect to the basis $\mathcal{E}$, .
(3) We denote by $\mathcal{E}^g = \{e^1, e^2, e^3\}$ and $\widetilde{\mathcal{B}}^g = \{\tilde{b}^1, \tilde{b}^2, \tilde{b}^3\}$ the reciprocal bases respectively of $\mathcal{E}$ and $\widetilde{\mathcal{B}}$. By Remark 3.13,

$$g(\tilde{b}^i, \tilde{b}_j) = \delta^i_j \qquad \text{and} \qquad g(e^i, e_j) = \delta^i_j$$

are independent of the choice of the basis. It follows that:
(a) $\delta^i_j = g(\tilde{b}^i, \tilde{b}_j) = [{}^{\mathrm{t}}\tilde{b}^i]_{\widetilde{\mathcal{B}}} G_{\widetilde{\mathcal{B}}} [\tilde{b}_j]_{\widetilde{\mathcal{B}}}$;
(b) $\delta^i_j = g(\tilde{b}^i, \tilde{b}_j) = [{}^{\mathrm{t}}\tilde{b}^i]_{\mathcal{E}} G_{\mathcal{E}} [\tilde{b}_j]_{\mathcal{E}}$;
(c) $\delta^i_j = g(e^i, e_j) = [{}^{\mathrm{t}}e^i]_{\widetilde{\mathcal{B}}} G_{\widetilde{\mathcal{B}}} [e_j]_{\widetilde{\mathcal{B}}}$;
(d) $\delta^i_j = g(e^i, e_j) = [{}^{\mathrm{t}}e^i]_{\mathcal{E}} G_{\mathcal{E}} [e_j]_{\mathcal{E}}$.

Using (1), (2) and the appropriate formula among (a), (b), (c) and (d), compute $[\tilde{b}^i]_{\widetilde{\mathcal{B}}}$, $[\tilde{b}^i]_{\mathcal{E}}$, $[e^i]_{\mathcal{E}}$ and $[e^i]_{\widetilde{\mathcal{B}}}$. For some of these, you will probably want to apply the same technique as in Example 3.19.

3.2.2. Properties of Reciprocal Bases.

CLAIM 3.21. Given a vector space V with a basis $\mathcal{B}$ and an inner product $g : V \times V \to \mathbb{R}$, a reciprocal basis *exists* and is *unique*.

As we pointed out right after the definition of reciprocal basis, what this claim really says is that there is a set of vectors $\{b^1, \dots, b^n\}$ in V that satisfy (3.9), that form a basis and that this basis is unique.

PROOF. Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be the given basis. Any other basis $\{b^1, \dots, b^n\}$ is related to $\mathcal{B}$ by the relation

$$b^i = M^{ij} b_j \tag{3.11}$$

for some *invertible* matrix M. We want to show that there exists a *unique* matrix M such that, when (3.11) is plugged into $g(b^i, b_j)$, we have

$$g(b^i, b_j) = \delta^i_j \,. \tag{3.12}$$

From (3.11) and (3.12) we obtain

$$\delta^i_j = g(b^i, b_j) = g(M^{ik} b_k, b_j) = M^{ik} g(b_k, b_j) = M^{ik} g_{kj} \,,$$

which, in matrix notation becomes

$$I = MG \,,$$

where G is the matrix of g with respect to $\mathcal{B}$ whose entries are g_{ij} as in (3.4). Since G is invertible because it is positive definite, then $M = G^{-1}$ exists and is unique. □

REMARK 3.22. Note that in the course of the proof we have found that, since $M = L_{\mathcal{B}^g\mathcal{B}}$, then

$$\boxed{G = (L_{\mathcal{B}^g\mathcal{B}})^{-1} = L_{\mathcal{B}\mathcal{B}^g}} \,.$$

We denote with g^{ij} the entries of $M = G^{-1}$. From the above discussion, it follows that with this notation

$$\boxed{g^{ik} g_{kj} = \delta^i_j} \tag{3.13}$$

as well as

$$\boxed{b^i = g^{ij} b_j} \,, \tag{3.14}$$

or[10]

$$\boxed{\begin{pmatrix} b^1 & \dots & b^n \end{pmatrix} = \begin{pmatrix} b_1 & \dots & b_n \end{pmatrix} G^{-1}} \,. \tag{3.15}$$

[10]Note that the following, like previously remarked, is a purely symbolic expression that has the only advantage of encoding the n expressions in (3.15).

(you need to understand why G^{-1} has to be multiplied on the right). Note that this is consistent with the findings in §1.3.2. We can now compute $g(b^i, b^j)$

$$\begin{aligned} g(b^i, b^j) &\overset{(3.14)}{=} g(g^{ik} b_k, g^{j\ell} b_\ell) = g^{ik} g^{j\ell} g(b_k, b_\ell) \\ &= g^{ik} g^{j\ell} g_{k\ell} \overset{(3.5)}{=} g^{ik} g^{j\ell} g_{\ell k} \overset{(3.13)}{=} g^{ik} \delta_k^j = g^{ij}, \end{aligned}$$

where we used in the second equality the bilinearity of g. Thus, similarly to (3.1), we have

$$\boxed{g^{ij} = g(b^i, b^j)}. \tag{3.16}$$

Exercise 3.23. In the setting of Exercise 3.20, verify (3.15) in the particular cases of $\mathcal{E}$ and $\mathcal{E}^g$ and of $\widetilde{\mathcal{B}}$ and $\widetilde{\mathcal{B}}^g$, that is verify that

(1) $\begin{pmatrix} e^1 & e^2 & e^3 \end{pmatrix} = \begin{pmatrix} e_1 & e_2 & e_3 \end{pmatrix} G_{\mathcal{E}}^{-1}$, and
(2) $\begin{pmatrix} \tilde{b}^1 & \tilde{b}_2 & \tilde{b}_3 \end{pmatrix} = \begin{pmatrix} \tilde{b}_1 & \tilde{b}_2 & \tilde{b}_3 \end{pmatrix} G_{\widetilde{\mathcal{B}}}^{-1}$.

Recall in fact that in (3.15), because of the way it was obtained, G is the matrix of g with respect to the basis $\mathcal{B}$.

Given that we just proved that reciprocal bases are unique, we can talk about *the* reciprocal basis (of a fixed vector space V associated to a basis and an inner product).

Claim 3.24. The reciprocal basis is **contravariant**.

Proof. Let $\mathcal{B}$ and $\widetilde{\mathcal{B}}$ be two bases of V and $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$ be the corresponding matrix of the change of basis, with $\Lambda = L^{-1}$. Recall that this means that

$$\tilde{b}_i = L_i^j b_j .$$

We have to check that if $\mathcal{B}^g = \{b^1, \dots, b^n\}$ is a reciprocal basis for $\mathcal{B}$, then the basis $\{\tilde{b}^1, \dots, \tilde{b}^n\}$ defined by

$$\tilde{b}^i = \Lambda_k^i b^k \tag{3.17}$$

is a reciprocal basis for $\widetilde{\mathcal{B}}$. Then the assertion will be proven, since $\{\tilde{b}^1, \dots, \tilde{b}^n\}$ is contravariant by construction.

To check that $\{\tilde{b}^1, \dots, \tilde{b}^n\}$ is the reciprocal basis, we need with check that with the choice of $\tilde{b}^i$ as in (3.17), the property (3.1) of the reciprocal basis is verified, namely that

$$g(\tilde{b}^i, \tilde{b}_j) = \delta_j^i .$$

But in fact,

$$g(\tilde{b}^i, \tilde{b}_j) \overset{(3.17)}{=} g(\Lambda_k^i b^k, L_j^\ell b_\ell) = \Lambda_k^i L_j^\ell g(b^k, b_\ell) \overset{(3.12)}{=} \Lambda_k^i L_j^\ell \delta_\ell^k = \Lambda_k^i L_j^k = \delta_j^i ,$$

where the second equality comes from the bilinearity of g, the third from the property (3.9) defining reciprocal basis and the last from the fact that $\Lambda = L^{-1}$. □

Suppose now that V is a vector space with a basis $\mathcal{B}$ and that $\mathcal{B}^g$ is the reciprocal basis of V with respect to $\mathcal{B}$ and to a fixed inner product $g : V \times V \to \mathbb{R}$. Then there are two ways of writing a vector $v \in V$, namely

$$v = \underbrace{v^i b_i}_{\text{with respect to } \mathcal{B}} = \underbrace{v_j b^j}_{\text{with respect to } \mathcal{B}^g} .$$

Recall that the (ordinary) coordinates of v with respect to $\mathcal{B}$ are *contravariant* (see Example 0.2).

CLAIM 3.25. Vector coordinates with respect to the reciprocal basis are **covariant**.

PROOF. This will follow from the fact that the reciprocal basis is contravariant and the idea of the proof is the same as in Claim 3.24.

Namely, let $\mathcal{B}, \widetilde{\mathcal{B}}$ be two bases of V, $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$ the matrix of the change of basis and $\Lambda = L^{-1}$. Let $\mathcal{B}^g$ and $\widetilde{\mathcal{B}}^g$ be the corresponding reciprocal bases and $v = v_j b^j$ a vector with respect to $\mathcal{B}^g$.

It is enough to check that the numbers

$$\tilde{v}_i := L_i^j v_j$$

are the coordinates of v with respect to $\widetilde{\mathcal{B}}^g$, because in fact these coordinates are covariant by definition. But in fact, using this and (3.17), we obtain

$$\tilde{v}_i \tilde{b}^i = (L_i^j v_j)(\Lambda_k^i b^k) = \underbrace{L_i^j \Lambda_k^i}_{\delta_k^j} v_j b^k = v_j b^j = v$$

□

DEFINITION 3.26. The coordinates v_i of a vector $v \in V$ with respect to the reciprocal basis $\mathcal{B}^g$ are called the **covariant coordinates** of v.

3.2.3. Change of Basis from a Basis $\mathcal{B}$ to its Reciprocal Basis $\mathcal{B}^g$.

We want to look now at the direct relationship between the covariant and the contravariant coordinates of a vector v. Recall that we can write

$$\underbrace{v^i b_i}_{\text{with respect to } \mathcal{B}} = v = \underbrace{v_j b^j}_{\text{with respect to } \mathcal{B}^g} .$$

from which we obtain

$$(v^i g_{ij}) b^j = v^i (g_{ij} b^j) = v^i b_i = v = v_j b^j ,$$

and hence, by comparing the coefficients of b^j,

$$\boxed{v_j = v^i g_{ij}} \quad \text{or} \quad \boxed{{}^{\mathrm{t}}[v]_{\mathcal{B}^g} = G[v]_{\mathcal{B}}}, . \tag{3.18}$$

Likewise, from

$$v^i b_i = v = v_j b^j = v_j (g^{ji} b_i) = (v_j g^{ji}) b_i ,$$

follows that

$$\boxed{v^i = v_j g^{ji}} \quad \text{or} \quad \boxed{[v]_{\mathcal{B}} = G^{-1\mathsf{t}}[v]_{\mathcal{B}^g}}. \tag{3.19}$$

EXAMPLE 3.27. Let $\mathcal{E} = \{e_1, e_2, e_3\}$ be the standard basis of $\mathbb{R}^3$ and let

$$G = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

be the matrix of g with respect to $\mathcal{E}$. In (3.10) Example 3.19, we saw that

$$\mathcal{E}^g = \left\{ b^1 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, b^2 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, b^3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is the reciprocal basis. We find the covariant coordinates of $v = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$ with respect to $\mathcal{E}^g$ using (3.18), namely

$$[v]_{\mathcal{E}^g} = G[v]_{\mathcal{E}} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 7 \end{pmatrix}.$$

The following computation double checks this result:

$$v_i b^i = (-1) \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} + 7 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

□

EXAMPLE 3.28. Let $V := \mathbb{R}[x]_1$ be the vector space of polynomials of degree ≤ 1 (that is, "linear" polynomials of the form $a + bx$). Let $g : V \times V \to \mathbb{R}$ be defined by

$$g(p, q) := \int_0^1 p(x)q(x)dx\,,$$

and let $\mathcal{B} := \{1, x\}$ be a basis of V. Determine:

(1) the matrix G;
(2) the matrix G^{-1};
(3) the reciprocal basis $\mathcal{B}^g$;
(4) the contravariant coordinates of $p(x) = 6x$ (that is the coordinates of $p(x)$ with respect to $\mathcal{B}$);
(5) the covariant coordinates of $p(x) = 6x$ (that is the coordinates of $p(x)$ with respect to $\mathcal{B}^g$).

	$\mathcal{B} = \{b_1, \ldots, b_n\}$ basis	$\mathcal{B}^g = \{b^1, \ldots, b^n\}$ reciprocal basis
these bases are related by	$g(b^i, b_j) = \delta^i_j$	
a vector v has now two sets of coordinates	$v = v^i b_i$ contravariant coordinates	$v = v_i b^i$ covariant coordinates
the matrices of g are	$g_{ij} = g(b_i, b_j)$	$g^{ij} = g(b^i, b^j)$
these matrices are inverse of each other, that is,	$g^{ik} g_{kj} = \delta^i_j$	
the basis and the reciprocal basis satisfy	$b_i = g_{ij} b^j$	$b^i = g^{ij} b_j$
covariant and contravariant coordinates are related by	$v^i = g^{ij} v_j$	$v_i = g_{ij} v^j$

TABLE 3. Summary of covariance and contravariance of vector coordinates

(1) The matrix G has entries $g_{ij} = g(b_i, b_i)$, that is

$$g_{11} = g(b_1, b_1) = \int_0^1 (b_1)^2 dx = \int_0^1 dx = 1$$

$$g_{12} = g(b_1, b_2) = \int_0^1 b_1 b_2 dx = \int_0^1 x = \frac{1}{2}$$

$$g_{21} = g(b_2, b_1) = \int_0^1 b_2 b_1 dx = \frac{1}{2}$$

$$g_{22} = g(b_2, b_2) = \int_0^1 (b_2)^2 dx = \int_0^1 x^2 dx = \frac{1}{3},$$

so that

$$G = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix} .$$

(2) Since $\det G = 1 \cdot \frac{1}{3} - \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{12}$, then by using (1.7), we get

$$G^{-1} = \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix} , .$$

(3) Using (3.15), we obtain that

$$\begin{pmatrix} b^1 & b^2 \end{pmatrix} = \begin{pmatrix} 1 & x \end{pmatrix} G^{-1} = \begin{pmatrix} 1 & x \end{pmatrix} \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix} = \begin{pmatrix} 4 - 6x & -6 + 12x \end{pmatrix} ,$$

so that $\mathcal{B}^g = \{4 - 6x, -6 + 12x\}$.

(4) $p(x) = 6x = 0 \cdot 1 + 6 \cdot x$, so that $p(x)$ has contravariant coordinates $[p(x)]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}$.

(5) From (3.18) it follows that if $v = p(x)$, then

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = G \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 0 \\ 6 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix} .$$

We can check this result:

$$v_1 b^1 + v_2 b^2 = 3 \cdot (4 - 6x) + 2 \cdot (-6 + 12x) = 6x .$$

□

3.2.4. Isomorphisms Between a Vector Space and its Dual.

We saw already in Proposition 2.9, that if V is a vector space and V^* is its dual, then $\dim V = \dim V^*$. In particular, this means that V and V^* can be identified, once we choose a basis $\mathcal{B}$ of V and a basis $\mathcal{B}^*$ of V^*. In fact, the basis $\mathcal{B}^*$ of V^* is given once we choose the basis $\mathcal{B}$ of V, as the dual basis of V^* with respect to $\mathcal{B}$. Then there is the following correspondence:

$$v \in V \leftrightsquigarrow \alpha \in V^* ,$$

exactly when v and α have the same coordinates, respectively with respect to $\mathcal{B}$ and $\mathcal{B}^*$. However, this correspondence depends on the choice of the basis $\mathcal{B}$ and hence *not canonical*.

If however V is endowed with an inner product, then there is a **canonical identification** of V with V^* that is, an identification that does not depend on the basis $\mathcal{B}$ of V. In fact, let $g : V \times V \to \mathbb{R}$ be an inner product and let $v \in V$. Then

$$\begin{aligned} g(v, \cdot\,) : V &\longrightarrow \mathbb{R} \\ w &\longmapsto g(v, w) \end{aligned}$$

is a linear form and hence we have the following *canonical* identification given by the metric

$$
\begin{aligned}
V &\longleftrightarrow \quad V^* \\
v &\longleftrightarrow v^* := g(v, \cdot\,)\,.
\end{aligned}
\tag{3.20}
$$

Note that the isomorphism[11] sends the zero vector to the linear form identically equal to zero, since $g(v, \cdot\,) \equiv 0$ if and only if $v = 0$ by positive definiteness of g.

So far, we have two bases of the vector space V, namely the basis $\mathcal{B}$ and the reciprocal basis $\mathcal{B}^g$ and we have also the dual basis of the dual vector space V^*. It turns out that, under the isomorphism (3.20), the reciprocal basis of V and the dual basis of V^* correspond to each other. This follows from the fact that, under the isomorphism (3.20) an element of the reciprocal basis b^i corresponds to the linear form $g(b^i, \cdot\,)$

$$b^i \longleftrightarrow g(b^i, \cdot\,)$$

and the linear form $g(b^i, \cdot\,) : V \to \mathbb{R}$ has the property that

$$g(b^i, b_j) = \delta^i_j\,.$$

We conclude that

$$g(b^i, \cdot\,) \equiv \beta^i\,,$$

so under the canonical identification between V and V^* the reciprocal basis of V corresponds to the dual basis of V^*.

3.2.5. Geometric Interpretation.

Let $g : V \times V \to \mathbb{R}$ be an inner product and $\mathcal{B} = \{b_1, \dots, b_n\}$ a basis of V with reciprocal basis $\mathcal{B}^g$. Let $v = v_i b^i \in V$ be a vector written in terms of its covariant coordinates (that is, the coordinates with respect to the reciprocal basis). Then

$$g(v, b_k) = g(v_i b^i, b_k) = v_i \underbrace{g(b^i, b_k)}_{\delta^i_k} = v_k\,,$$

so that (3.8) becomes

$$\operatorname{proj}_{b_k} v = \frac{v_k}{g(b_k, b_k)} b_k\,.$$

If we assume that the elements of the basis $\mathcal{B} = \{b_1, \dots, b_n\}$ are unit vectors, then (3.8) further simplifies to give

$$\operatorname{proj}_{b_k} v = v_k b_k\,. \tag{3.21}$$

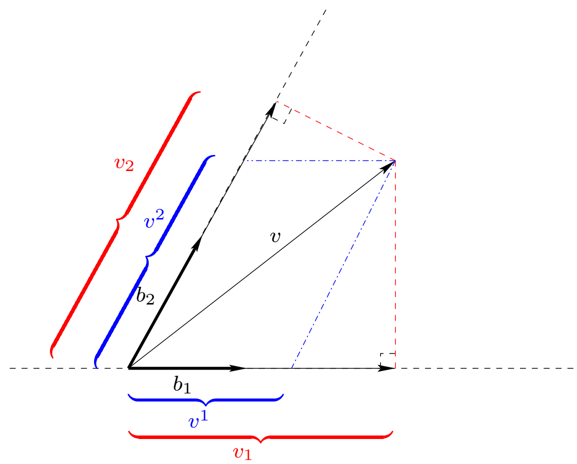
This equation shows the following:

Fact 3.29. *The covariant coordinates of v give the orthogonal projection of v onto $b_1, \dots, b_n$.*

[11]Recall that an *isomomorphism* $T : V \to W$ is an invertible linear transformation.

Likewise, the following holds basically by definition:

FACT 3.30. *The contravariant coordinates of v give the "parallel" projection of v onto $b_1, \ldots, b_n$.*



3.3. Relevance of Covariance and Contravariance

Why do we need or care for covariant and contravariant components?

3.3.1. Physical Relevance.

Consider the following physical problem: Calculate the work performed by a force F on a particle to move the particle by a small displacement dx, in the Euclidean plane. The work performed should be independent of the choice of the coordinate system (i.e choice of basis) used. For the work to remain independent of choice of basis we will see that, if the components of the displacement change contravariantly, then the components of the force should change covariantly.

To see this let $\mathcal{B} = \{b_1, b_2\}$ be a basis of the Euclidean plane. Suppose the force $F = (F_1, F_2)$ is exerted on a particle that moves with a displacement $dx = (dx^1, dx^2)$. Then the work done is given by

$$dW = F_1 dx^1 + F_2 dx^2 .$$

Suppose we are given another coordinate system $\widetilde{\mathcal{B}} := \{\tilde{b}_1, \tilde{b}_2\}$ and let $F = (\tilde{F}_1, \tilde{F}_2)$ and $dx = (d\tilde{x}^1, d\tilde{x}^2)$. Then

$$dW = \tilde{F}_1 d\tilde{x}^1 + \tilde{F}_2 d\tilde{x}^2 .$$

Now assume that the coordinates of dx change contravariantly;

$$d\tilde{x}^i = \Lambda^i_j dx^j ,$$

or, equivalently,

$$dx^i = L^i_j d\tilde{x}^j ,$$

where $\Lambda = L^{-1}$ and $L = (L^i_j)$ is the change of basis matrix from $\mathcal{B}$ to $\widetilde{\mathcal{B}}$.

$$\begin{aligned} dW =& F_1 dx^1 + F_2 dx^2 \\ =& F_1[L^1_1 d\tilde{x}^1 + L^1_2 d\tilde{x}^2] + F_2[L^2_1 d\tilde{x}^1 + L^2_2 d\tilde{x}^2] \\ =& (F_1 L^1_1 + F_2 L^2_1) d\tilde{x}^1 + (F_1 L^1_2 + F_2 L^2_2) d\tilde{x}^2 \end{aligned}$$

Since the work performed is independent of basis chosen we also have

$$dW = \tilde{F}_1 d\tilde{x}^1 + \tilde{F}_2 d\tilde{x}^2 .$$

This gives that

$$\begin{aligned} \tilde{F}_1 &= F_1 L^1_1 + F_2 L^2_1 \\ \tilde{F}_2 &= F_1 L^1_2 + F_2 L^2_2 . \end{aligned}$$

Hence the coordinates of F transform covariantly; $\tilde{F}_i = L^j_i F_j$. Using matrices this can be written as $(\tilde{F}_1, \tilde{F}_2) = (F_1, F_2)L$.

3.3.2. Distinction Vanishes when Restricting to Orthonormal Bases.

In the presence of an inner product $g : V \times V \to \mathbb{R}$ *and* if we restrict ourselves to orthonormal bases, the distinction between covariance and contravariance vanishes! In fact, it all becomes a matter of transposition: writing vectors as columns or as rows.

Why is that?

First, as we saw, the reciprocal of an orthonormal basis is equal to itself, so covariant and contravariant coordinates are equal in this special case.

Moreover, when we change from one orthonormal basis to another by a change of basis matrix L, the inverse change is given simply by the transpose of L. Here is a proof.

Let $\mathcal{B} = \{b_1, \dots, b_n\}$ and $\widetilde{\mathcal{B}} = \{\tilde{b}_1, \dots, \tilde{b}_n\}$ be two *orthonormal* bases of V and let $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$ be the corresponding matrix of the change of basis. This means, as always, that

$$\tilde{b}_i = L^j_i b_j .$$

Since $\mathcal{B}$ and $\widetilde{\mathcal{B}}$ are orthonormal, we have $g(b_i, b_j) = \delta_{ij} = g(\tilde{b}_i, \tilde{b}_j)$. Therefore, we have

$$\delta_{ij} = g(\tilde{b}_i, \tilde{b}_j) = g(L^k_i b_k, L^\ell_j b_\ell) = L^k_i L^\ell_j g(b_k, b_\ell) = L^k_i L^\ell_j \delta_{k\ell} == L^k_i L^k_j ,$$

showing that $L^{\mathsf{t}} L = I$, that is, $L^{-1} = {}^{\mathsf{t}}L$. Such a matrix L is called an *orthogonal matrix*.

We conclude that, in this special case, we see no distinction between covariant and contravariant behaviour, it is simply a matter of transposition.

CHAPTER 4

Tensors

4.1. Towards General Tensors

Let V be a vector space. Up to now, we saw several objects related to V, which we called *tensors*. We summarize them in Table 4. From these, we infer the definition of a tensor of type $(0, q)$ for all $q \in \mathbb{N}$, but we cannot say the same for a tensor of type $(p, 0)$ for all $p \in \mathbb{N}$, even less of general type (p, q). The next discussion will lead us to tensors of type $(p, 0)$, and in the meantime we will discuss an important isomorphism.

4.1.1. Canonical Isomorphism between V and $(V^*)^*$.

We saw in §3.2.4 that any vector space is isomorphic to its dual, though in general the isomorphism is *not* canonical, that is, it depends on the choice of a basis. We also saw that, if there is an inner product on V, then there is a canonical isomorphism. The point of this section is to show that, even without an inner product, there is always a *canonical* isomorphism between V and its **bidual** $(V^*)^*$, that is the dual of its dual.

To see this, let us observe first of all that

$$\dim V = \dim(V^*)^* . \tag{4.1}$$

Tensor	Components	Behavior under a change of basis	Type
vector in V	v^i	contravariant tensor	$(1, 0)$
linear form $V \to \mathbb{R}$	α_j	covariant tensor	$(0, 1)$
linear transformation $V \to V$	A^i_j	tensor of mixed character: contravariant and covariant	$(1, 1)$
bilinear form $V \times V \to \mathbb{R}$	B_{ij}	covariant 2-tensor	$(0, 2)$
k-linear form $V \times \cdots \times V \to \mathbb{R}$	$F_{i_1 i_2 \dots i_k}$	covariant k-tensor	$(0, k)$

TABLE 4. Covariance and contravariance of earlier tensors

In fact, for any vector space W, we saw in Proposition 2.9 that $\dim W = \dim W^*$. If we apply this equality both to $W = V$ and to $W = V^*$, we obtain

$$\dim V = \dim V^* \quad \text{and} \quad \dim V^* = \dim(V^*)^*,$$

from which (4.1) follows immediately. We deduce (for instance, using §3.2.4) that V and $(V^*)^*$ are isomorphic, and we only have to see that there is a *canonical* isomorphism.

To this end, observe that a vector $v \in V$ gives rise to a linear form on V^* defined by

$$\begin{aligned} \varphi_v : V^* &\longrightarrow \mathbb{R} \\ \alpha &\longmapsto \alpha(v). \end{aligned}$$

Then we can define a linear map as follows:

$$\begin{aligned} \Phi : V &\longrightarrow (V^*)^* \\ v &\longmapsto \varphi_v. \end{aligned} \tag{4.2}$$

Since, for any linear map $T : V \to W$ between vector spaces, we have the dimension formula (known as the *rank-nullity theorem in Linear Algebra*):

$$\dim V = \dim \operatorname{im}(T) + \dim \ker(T),$$

it will be enough to show that $\ker \Phi = \{0\}$, because then

$$\dim V = \dim \operatorname{im}(\Phi) = \dim(V^*)^*,$$

and we can conclude that $\operatorname{im}(\Phi) = (V^*)^*$, hence Φ is an isomorphism. Notice that we have *not* chosen a basis to define the isomorphism Φ.

To see that $\ker \Phi = \{0\}$, observe that this kernel consists of all vectors $v \in V$ such that $\alpha(v) = 0$ for all $\alpha \in V^*$. We want to see that the only vector $v \in V$ for which this happens is the zero vector. In fact, if $v \in V$ is nonzero and $\mathcal{B} = \{b_1, \dots, b_n\}$ is *any* basis of V, then we can write $v = v^i b_i$, where at least one coordinate, say v^j, is not zero. In that case, if $\mathcal{B}^* = \{\beta_1, \dots, \beta_n\}$ is the dual basis, we have $\beta_j(v) = v^j \neq 0$, thus we have found an element in V^* not vanishing on this v. We record this fact as follows:

FACT 4.1. *Let V be a vector space and V^* its dual. The dual $(V^*)^*$ of V^* is canonically isomorphic to V. The canonical isomorphism $\Phi : V \to (V^*)^*$ takes $v \in V$ to the linear form on the dual $\varphi_v : V^* \to \mathbb{R}$, $\varphi_v(\alpha) := \alpha(v)$.*

4.1.2. $(2,0)$-Tensors.

Recall that the dual of a vector space V is the vector space

$$V^* := \{\text{linear forms } \alpha : V \to \mathbb{R}\} = \{(0,1)\text{-tensors}\}.$$

Taking now the dual of the vector space V^*, we obtain

$$(V^*)^* := \{\text{linear forms } \varphi : V^* \to \mathbb{R}\}.$$

Using the canonical isomorphism $\Phi : V \to (V^*)^*$ and the fact that coordinate vectors are contravariant, we conclude that

$$\{\text{linear forms } \varphi : V^* \to \mathbb{R}\} = (V^*)^* \cong V = \{(1,0)\text{-tensors}\}\,.$$

So, changing the vector space from V to its dual V^* seems to have had the effect of converting covariant tensors of type $(0,1)$ into contravariant ones of type $(1,0)$.

We are going to apply the above principle to convert covariant tensors of type $(0,2)$ into contravariant ones of type $(2,0)$. Recall that

$$\{(0,2)\text{-tensors}\} = \{\text{bilinear maps } \varphi : V \times V \to \mathbb{R}\}$$

and consider now

$$\{\text{bilinear maps } \sigma : V^* \times V^* \to \mathbb{R}\}\,.$$

Anticipating the contravariant character of such bilinear maps (to be proven in §4.1.4), we advance the following definition:

Definition 4.2. A **tensor of type** $(2,0)$ is a bilinear form on V^*, that is, a bilinear function $\sigma : V^* \times V^* \to \mathbb{R}$.

Then we have

$$\{(2,0)\text{-tensors}\} = \{\text{bilinear maps } \sigma : V^* \times V^* \to \mathbb{R}\}$$

and we denote this set $\mathrm{Bil}(V^* \times V^*, \mathbb{R})$.

Exercise 4.3. Check that $\mathrm{Bil}(V^* \times V^*, \mathbb{R})$ is a vector space. Just like in the case of $\mathrm{Bil}(V \times V, \mathbb{R})$ (cf. Exercise 2.25), it is enough to show that the zero map is in $\mathrm{Bil}(V^* \times V^*, \mathbb{R})$ and that if $\sigma, \tau \in \mathrm{Bil}(V^* \times V^*, \mathbb{R})$ and $c, d \in \mathbb{R}$, then the linear combination $c\sigma + d\tau$ is also in $\mathrm{Bil}(V^* \times V^*, \mathbb{R})$.

4.1.3. Tensor Product of Two Linear Forms on V^*.

If $v, w \in V$ are two vectors (i.e., are two $(1,0)$-tensors), we define

$$\sigma_{v,w} : V^* \times V^* \to \mathbb{R}$$

by

$$\sigma_{v,w}(\alpha, \beta) := \alpha(v)\beta(w)\,,$$

for any two linear forms $\alpha, \beta \in V^*$. Then $\sigma_{v,w}$ is indeed *bilinear*, i.e., linear in each variable, α and β, so is indeed a $(2,0)$-**tensor**. We denote

$$\boxed{\sigma_{v,w} =: v \otimes w}$$

and call this the **tensor product** of v and w.

NOTE 4.4. In general, we have

$$\boxed{v \otimes w \neq w \otimes v},$$

as there can be linear forms α, β such that $\alpha(v)\beta(w) \neq \alpha(w)\beta(v)$.

Similar to what we saw in §2.2.3, we find a basis for the space of $(2,0)$-tensors by considering the $(2,0)$-tensors defined by $b_i \otimes b_j$, where $\mathcal{B} = \{b_1, \dots, b_n\}$ is a basis of V.

PROPOSITION 4.5. *The elements* $b_j \otimes b_j$, $i, j = 1, \dots, n$ *form a basis of* $\mathrm{Bil}(V^* \times V^*, \mathbb{R})$. *Thus* $\dim \mathrm{Bil}(V^* \times V^*, \mathbb{R}) = n^2$.

The proof of this proposition is analogous to the one of Proposition 2.26 and we will not write it here. However, we state the crucial remark for the proof, analogous to Remark 2.27.

REMARK 4.6. As for linear forms and bilinear forms on V, in order to verify that two bilinear forms on V^* are the same, it is enough to verify that they are the same on every pair of elements of a basis of V^*. In fact, let σ, τ be two bilinear forms on V^*, let $\{\gamma^1, \dots, \gamma^n\}$ be a basis of V^* and let us assume that

$$\sigma(\gamma^i, \gamma^j) = \tau(\gamma^i, \gamma^j)$$

for all $1 \leq i, j, \leq n$. Let $\alpha = \alpha_i \gamma^i$ and $\beta = \beta_j \gamma^j$ be arbitrary elements of V^*. We now verify that $\sigma(\alpha, \beta) = \tau(\alpha, \beta)$. Because of the linearity in each variable, we have

$$\sigma(\alpha, \beta) = \sigma(\alpha_i \gamma^i, \beta_j \gamma^j) = \alpha_i \beta_j \sigma(\gamma^i, \gamma^j) = \alpha_i \beta_j \tau(\gamma^i, \gamma^j) = \tau(\alpha_i \gamma^i, \beta_j \gamma^j) = \tau(\alpha, \beta)\,.$$

□

4.1.4. Contravariance of $(2,0)$-Tensors.

Let $\sigma : V^* \times V^* \to \mathbb{R}$ be a bilinear form on V^*, that is, a so-called $(2,0)$-tensor. We want to verify that it behaves as we expect with respect to a change of basis. After choosing a basis $\mathcal{B} = \{b_1, \dots, b_n\}$ of V, we have the dual basis $\mathcal{B}^* = \{\beta^1, \dots, \beta^n\}$ of V^* and the basis $\{b_i \otimes b_j : i, j = 1, \dots, n\}$ of the space of $(2,0)$-tensors.

The $(2,0)$-tensor σ is represented by its components

$$\boxed{S^{ij} = \sigma(\beta^i, \beta^j)},$$

in the sense that

$$\boxed{\sigma = S^{ij} b_j \otimes b_j},$$

and the components S^{ij} can be arranged into a matrix[12]

$$S = \begin{pmatrix} S^{11} & \dots & S^{1n} \\ \vdots & \ddots & \vdots \\ S^{n1} & \dots & S^{nn} \end{pmatrix}$$

[12]Once again, contrary to the change of basis matrix L, here we have only upper indices. This reflects the contravariance of the underlying tensor, σ.

called the **matrix of the $(2,0)$-tensor** σ with respect to the chosen basis of V.

We look now at how the components of a $(2,0)$-tensor change with a change of basis. Let $\mathcal{B} = \{b_1,\dots,b_n\}$ and $\widetilde{\mathcal{B}} = \{\tilde{b}_1,\dots,\tilde{b}_n\}$ be two basis of V and let $\mathcal{B}^* := \{\beta^1,\dots,\beta^n\}$ and $\widetilde{\mathcal{B}}^* := \{\tilde{\beta}^1,\dots,\tilde{\beta}^n\}$ be the corresponding dual bases of V^*. Let $\sigma : V^* \times V^* \to \mathbb{R}$ be a $(2,0)$-tensor with components

$$S^{ij} = \sigma(\beta^i,\beta^j) \quad \text{and} \quad \widetilde{S}^{ij} = \sigma(\tilde{\beta}^i,\tilde{\beta}^j)$$

with respect to $\mathcal{B}^*$ and $\widetilde{\mathcal{B}}^*$, respectively. Let $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$ be the matrix of the change of basis from $\mathcal{B}$ to $\widetilde{\mathcal{B}}$, and let $\Lambda := L^{-1}$. Then, as seen in (1.4) and (2.11), we have that

$$\tilde{b}_j = L^i_j b_i \quad \text{and} \quad \tilde{\beta}^i = \Lambda^i_j \beta^j \,.$$

It follows that

$$\widetilde{S}^{ij} = \sigma(\tilde{\beta}^i,\tilde{\beta}^j) = \sigma(\Lambda^i_k \beta^k, \Lambda^j_\ell \beta^\ell) = \Lambda^i_k \Lambda^j_\ell \sigma(\beta^k,\beta^\ell) = \Lambda^i_k \Lambda^j_\ell S^{k\ell} \,,$$

where the first and the last equalities follow from the definition of $\widetilde{S}^{ij}$ and of $S^{k\ell}$, respectively, the second from the change of basis and the third from the bilinearity of σ. We conclude that

$$\boxed{\widetilde{S}^{ij} = \Lambda^i_k \Lambda^j_\ell S^{k\ell}} \,. \tag{4.3}$$

Hence, the bilinear form σ is a **contravariant** 2-tensor.

EXERCISE 4.7. Verify that, in terms of matrices (4.3) translates into

$$\boxed{\widetilde{S} = \Lambda S \Lambda^{\mathrm{t}}} \,.$$

Compare with (2.15).

4.2. Tensors of Type (p,q)

In general, a (p,q)-tensor (with $p,q = 0,1,2,\dots$) is defined to be a real-valued function of p covectors and of q vectors, which is linear in each of its arguments:

DEFINITION 4.8. A **tensor of type** (p,q) or (p,q)-**tensor** is a multilinear form (or $(p+q)$-linear function)

$$T : \underbrace{V^* \times \dots V^*}_{p} \times \underbrace{V \times \dots \times V}_{q} \longrightarrow \mathbb{R} \,.$$

By convention, a tensor of type $(0,0)$ is a real number, a.k.a. *scalar* (a constant funtion of no arguments).

The **order** of a tensor is the number of arguments that it takes: a tensor of type (p,q) has, thus, order $p+q$.[13]

[13]The *order* of a tensor is sometimes also called *rank*. However, *rank* of a tensor is often reserved for another notion closer to the notion of *rank* of a matrix and related to *decomposability* of tensors (see §4.3.1 and §4.3.2).

Earlier tensor	Viewed as multilinear function	Type
vector $v \in V$	$V^* \longrightarrow \mathbb{R}$ $\beta \longmapsto \beta(v)$	$(1, 0)$
linear form $\alpha \in V^*$	$V \longrightarrow \mathbb{R}$ $w \longmapsto \alpha(w)$	$(0, 1)$
linear transformation $F : V \to V$	$V^* \times V \longrightarrow \mathbb{R}$ $(\beta, w) \longmapsto \beta(F(w))$	$(1, 1)$
bilinear form $\varphi \in \mathrm{Bil}(V \times V, \mathbb{R})$	$V \times V \longrightarrow \mathbb{R}$ $(v, w) \longmapsto \varphi(v, w)$	$(0, 2)$
k-linear form φ on V	$V \times \cdots \times V \longrightarrow \mathbb{R}$ $(v_1, \dots, v_k) \longmapsto \varphi(v_1, \dots, v_k)$	$(0, k)$
bilinear form on V^* $\sigma \in \mathrm{Bil}(V^* \times V^*, \mathbb{R})$	$V^* \times V^* \longrightarrow \mathbb{R}$ $(\alpha, \beta) \longmapsto \sigma(\alpha, \beta)$	$(2, 0)$

TABLE 5. Earlier tensors viewed within general definition

If all the arguments of a tensor are vectors, i.e. $p = 0$, the tensor is said to be (purely) **covariant**. If the arguments are all linear forms, i.e. $q = 0$, the tensor is said to be (purely) **contravariant**. Otherwise, a (p, q)-tensor is of mixed character, p being its order of contravariance and q its order of covariance. Purely covariant tensors are what we earlier called multilinear forms. Purely contravariant tensors are sometimes called *polyadics*.[14]

Table 5 gives an overview of how the earlier examples of tensors fit in the above general definition.

Let T be a (p, q)-tensor, $\mathcal{B} = \{b_1, \dots, b_n\}$ a basis of V and $\mathcal{B}^* = \{\beta^1, \dots, \beta^n\}$ the corresponding dual basis of V^*. The **components** of T with respect to these bases are

$$\boxed{T^{i_1,\dots,i_p}_{j_1,\dots,j_q} := T(\beta^{i_1}, \dots, \beta^{i_p}, b_{j_1}, \dots, b_{j_q})}\,.$$

If, moreover, $\widetilde{\mathcal{B}} = \{\tilde{b}_1, \dots, \tilde{b}_n\}$ is another basis, $\widetilde{\mathcal{B}}^* = \{\tilde{\beta}^1, \dots, \tilde{\beta}^n\}$ the corresponding dual basis of V^* and $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$ the matrix of the change of basis with inverse $\Lambda := L^{-1}$, then the components of T with respect to these new bases are

$$\boxed{\widetilde{T}^{i_1,\dots,i_p}_{j_1,\dots,j_q} = \Lambda^{i_1}_{k_1} \dots \Lambda^{i_p}_{k_p} L^{\ell_1}_{j_1} \dots L^{\ell_q}_{j_q} T^{k_1,\dots,k_p}_{\ell_1,\dots,\ell_q}}\,.$$

The above formula displays the p-fold contravariant character and the q-fold covariant character of T.

[14]Whereas Latin roots are used for covariant tensors, like in *bi*linear form, Greek roots are used for contravariant tensors, like in *dy*adic.

The set of all tensors of type (p,q) on a vector space V with the natural operations of addition and scalar multiplication on tensors is itself a vector space denoted by

$$\boxed{\mathcal{T}^p_q(V) := \{\text{all } (p,q)\text{-tensors on } V\}}\,.$$

4.3. Tensor Product

We saw already in §2.2.3 and §2.3.3 the tensor product of two multilinear forms. Since multilinear forms are covariant tensors, we said that this corresponds to the tensor product of two covariant tensors. We can now define the tensor product of two tensors in general. This will further lead us to the tensor product of vector spaces.

4.3.1. Tensor Product for Tensors.

DEFINITION 4.9. Let

$$T : \underbrace{V^* \times \cdots \times V^*}_{p} \times \underbrace{V \times \cdots \times V}_{q} \longrightarrow \mathbb{R}$$

be a (p,q)-tensor and

$$U : \underbrace{V^* \times \cdots \times V^*}_{k} \times \underbrace{V \times \cdots \times V}_{\ell} \longrightarrow \mathbb{R}$$

a (k,ℓ) tensor. The **tensor product** $T \otimes U$ of T and U is the $(p+k, q+\ell)$-tensor

$$T \otimes U : \underbrace{V^* \times \cdots \times V^*}_{p+k} \times \underbrace{V \times \cdots \times V}_{q+\ell} \longrightarrow \mathbb{R}$$

defined by

$$\boxed{\begin{gathered}(T \otimes U)(\alpha_1, \dots, \alpha_{p+k}, v_1, \dots, v_{q+\ell}) \ := \\ T(\alpha_1, \dots, \alpha_p, v_1, \dots, v_q)U(\alpha_{p+1}, \dots, \alpha_{p+k}, v_{q+1}, \dots, v_{q+\ell})\,.\end{gathered}}$$

Although both $T \otimes U$ and $U \otimes T$ are tensors of the same type, in general we have

$$T \otimes U \neq U \otimes T\,.$$

So we say that the tensor product is *not commutative*. On the other hand, the tensor product is *associative*, since we always have

$$(S \otimes T) \otimes U = S \otimes (U \otimes T)\,.$$

Analogously to how we proceeded in the case of $(0,2)$-tensors, we compute the dimension of the vector space $\mathcal{T}^p_q(V)$. Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis of V and $\mathcal{B}^* := \{\beta^1, \dots, \beta^n\}$ the corresponding dual basis of V^*. Just like we saw in Proposition 4.5 in the case of $(0,2)$-tensors, we form a basis of $\mathcal{T}^p_q(V)$ by collecting all elements of the form

$$b_{i_1} \otimes b_{i_2} \otimes \cdots \otimes b_{i_p} \otimes \beta^{j_1} \otimes \beta^{j_2} \otimes \cdots \otimes \beta^{j_q}$$

where the indices $i_1, \ldots, i_p$ and $j_1, \ldots, j_q$ take all values between 1 and n. Since there are $n^p n^q = n^{p+q}$ elements in the above basis (corresponding to all possible choices of b_{i_k} and β^{j_ℓ}), we deduce that

$$\boxed{\dim \mathcal{T}^p_q(V) = n^{p+q}}.$$

If T is a (p,q)-tensor with components

$$T^{i_1,\ldots,i_p}_{j_1,\ldots,j_q} := T(\beta^{i_1}, \ldots, \beta^{i_p}, b_{j_1}, \ldots, b_{j_q}),$$

then we have, with a $(p+q)$-fold application of the Einstein convention, that

$$T = T^{i_1,\ldots,i_p}_{j_1,\ldots,j_q} b_{i_1} \otimes b_{i_2} \otimes \cdots \otimes b_{i_p} \otimes \beta^{j_1} \otimes \beta^{j_2} \otimes \cdots \otimes \beta^{j_q}.$$

A **simple tensor** (also called a **tensor of rank 1** or **pure tensor** or **decomposable tensor** or **elementary tensor**) of type (p,q) is a tensor T that can be written as a tensor product of the form

$$T = \underbrace{a \otimes b \otimes \ldots}_{p} \otimes \underbrace{\alpha \otimes \beta \otimes \ldots}_{q}$$

where $a, b, \ldots \in V$ and $\alpha, \beta, \ldots \in V^*$. The **rank of a tensor** T is then the minimum number of simple tensors that sum to T.[15] By convention, the zero tensor has rank 0 and a nonzero tensor of order zero, i.e., a nonzero scalar, has rank 1. A nonzero tensor of order 1 always has rank 1. Already among tensors of order 2 (and when $\dim V \geq 2$) there are tensors of rank greater than 1. Example 4.11 provides such an instance.

4.3.2. Tensor Product for Vector Spaces.

To complement the previous exposition and justify the notation $V^* \otimes V^*$ for the vector space of all bilinear forms on V (cf. §2.2.3), we aim in this section to give an idea of what the *tensor product* for finite-dimensional vector spaces should mean and of how the tensor product *for vector spaces* relates to the tensor product *for tensors*.

Let V and W be two vector spaces with $\dim V = n$ and $\dim W = m$. Choose $\{b_1, \ldots, b_n\}$ a basis of V and $\{a_1, \ldots, a_m\}$ a basis of W.[16]

Definition 4.10. The **tensor product** of V and W is the $(n \cdot m)$-dimensional vector space $V \otimes W$ with basis

$$\{b_i \otimes a_j : 1 \leq i \leq n, 1 \leq j \leq m\}.$$

[15] This notion of *rank* of a tensor extends the notion of *rank* of a matrix, as can be seen by considering tensors of order two and their corresponding matrices of components.

[16] There is a way of defining the tensor product for vector spaces without involving bases, but we will not do it here. That other, more abstract way shows elegantly that the tensor product of vector spaces does not depend on the choice of bases.

Elements of $V \otimes W$ are naturally referred to as *tensors*; further below we address the connection to the previous notion of tensor. By definition, they are linear combinations of the $b_i \otimes a_j$. By viewing $b_i \in V$ as a *linear map* (cf. Table 5)

$$b_i : V^* \longrightarrow \mathbb{R}, \qquad \beta \longmapsto \beta(b_i),$$

and similarly for $a_j \in W$ as

$$a_j : W^* \longrightarrow \mathbb{R}, \qquad \alpha \longmapsto \alpha(a_j),$$

we may regard the symbol $b_i \otimes a_j$ as a bilinear map

$$V^* \times W^* \longrightarrow \mathbb{R}, \qquad (\beta, \alpha) \longmapsto \beta(b_i)\alpha(a_j).$$

The tensor product $V \otimes W$ is endowed with a bilinear map from the cartesian product $V \times W$

$$\Psi \ : \ V \times W \ \longrightarrow \ V \otimes W$$

defined as follows. If

$$v = v^i b_i \in V \qquad \text{and} \qquad w = w^j a_j \in W,$$

then $\Psi(v, w) =: v \otimes w$ is the element of $V \otimes W$ with coordinates $v^i w^j$ with respect to the basis $\{b_i \otimes a_j : 1 \le i \le n, 1 \le j \le m\}$, so that the following holds:

$$v \otimes w = \left(v^i b_i\right) \otimes \left(w^j a_j\right) = \left(v^i w^j\right) b_i \otimes a_j .$$

Notice that the ranges for the indices are different: $1 \le i \le n$, $1 \le j \le m$. The numbers $v^i w^j$ may be viewed as obtained by the so-called *outer product* of the coordinate vectors of v and w yielding an $n \times m$ matrix:

$$[v]_{\mathcal{B}}{}^{\mathsf{t}}[w]_{\mathcal{A}} = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \begin{pmatrix} w^1 & \dots & w^m \end{pmatrix} = \begin{pmatrix} v^1 w^1 \dots v^1 w^m \\ \vdots \qquad \vdots \\ v^n w^1 \dots v^n w^m \end{pmatrix} .$$

An element of $V \otimes W$ that is in the image of the map Ψ, that is, an element of $V \otimes W$ that can be written as $v \otimes w$ for some $v \in V$ and $w \in W$ is called a **simple tensor** (or **tensor of rank 1** or **pure tensor** or **decomposable tensor** or **elementary tensor**). Yet keep in mind, that the map Ψ is usually by far *not* surjective. In particular, one can show that, if $v_1, v_2 \in V$ are linearly independent and $w_1, w_2 \in W$ are also linearly independent, then the sum $v_1 \otimes w_1 + v_2 \otimes w_2$ is not a pure tensor. Checking the first instance of this phenomenon is left as the next exercise. The proof in general goes along somewhat similar lines, but gets sophisticated.[17]

EXERCISE 4.11. Check that if both V and W are 2-dimensional with respective bases $\{b_1, b_2\}$ and $\{a_1, a_2\}$, then $b_1 \otimes a_1 + b_2 \otimes a_2$ is not a pure tensor.

The following proposition gives some useful identifications.

[17] If you are interested in learning more, look up the *Segre embedding* from Algebraic Geometry.

PROPOSITION 4.12. *Let V and W be vector spaces with $\dim V = n$ and $\dim W = m$ and let*

$$\mathrm{Lin}(V, W^*) := \{\textit{linear maps } V \to W^*\}\,.$$

Then

$$\begin{aligned}\mathrm{Bil}(V \times W, \mathbb{R}) &\cong \mathrm{Lin}(V, W^*)\\ &\cong \mathrm{Lin}(W, V^*)\\ &\cong V^* \otimes W^*\\ &\cong (V \otimes W)^*\\ &= \mathrm{Lin}(V \otimes W, \mathbb{R})\,.\end{aligned}$$

PROOF. Here is the idea behind this chain of identifications. Let $f \in \mathrm{Bil}(V \times W, \mathbb{R})$, that is, $f : V \times W \to \mathbb{R}$ is a bilinear function, in particular it takes two vectors, $v \in V$ and $w \in W$, as input and gives a real number $f(v, w) \in \mathbb{R}$ as output. If, however, we only feed f one vector $v \in V$ as input, then there is a remaining spot waiting for a vector $w \in W$ to produce a real number. Since f is linear in V and in W, the map $f(v, \cdot\,) : W \to \mathbb{R}$ is a linear form, so $f(v, \cdot\,) \in W^*$, hence f gives us an element in $\mathrm{Lin}(V, W^*)$. There is then a linear map

$$\begin{aligned}\mathrm{Bil}(V \times W, \mathbb{R}) &\longrightarrow \mathrm{Lin}(V, W^*)\\ f \qquad &\longmapsto \qquad T_f\,,\end{aligned}$$

where

$$T_f(v)(w) := f(v, w)\,.$$

Conversely, any $T \in \mathrm{Lin}(V, W^*)$ can be identified with a bilinear map $f_T \in \mathrm{Bil}(V \times W, \mathbb{R})$ defined by

$$f_T(v, w) := T(v)(w)\,.$$

Since $f_{T_f} = f$ and $T_{f_T} = T$, we have proven the first identification in the proposition.

Analogously, if the input is only a vector $w \in W$, then $f(\,\cdot\,, w) : V \to \mathbb{R}$ is a linear map and we now see that $f \in \mathrm{Bil}(V \times W, \mathbb{R})$ defines a linear map $U_f \in \mathrm{Lin}(W, V^*)$. The same reasoning as in the previous paragraph, shows that $\mathrm{Bil}(V \times W, \mathbb{R}) \cong \mathrm{Lin}(W, V^*)$.

To proceed with the identifications, observe that, because of our definition of $V^* \otimes W^*$, we have

$$\mathrm{Bil}(V \times W, \mathbb{R}) \cong V^* \otimes W^*\,,$$

since these spaces both have basis

$$\{\beta^i \otimes \alpha^j : 1 \le i \le n,\, 1 \le j \le m\}\,,$$

where $\{b_1, \dots, b_n\}$ is a basis of V with corresponding dual basis $\{\beta^1, \dots, \beta^n\}$ of V^*, and $\{a_1, \dots, a_n\}$ is a basis of W with corresponding dual basis $\{\alpha^1, \dots, \alpha^n\}$ of W^*.

Finally, an element $D_{ij}\beta^i\otimes\alpha^j \in V^*\otimes W^*$ may be viewed as a linear map $V\otimes W \to \mathbb{R}$, that is as an element of $(V\otimes W)^*$ by

$$\begin{aligned} V\otimes W &\longrightarrow \qquad\qquad \mathbb{R} \\ C^{k\ell}b_k\otimes a_\ell &\longmapsto D_{ij}C^{k\ell}\underbrace{\beta^i(b_k)}_{\delta^i_k}\underbrace{\alpha^j(a_\ell)}_{\delta^j_\ell} = D_{ij}C^{k\ell}\,. \end{aligned}$$

□

Because of the identification $\mathrm{Bil}(V\times W,\mathbb{R})\cong\mathrm{Lin}(V\otimes W,\mathbb{R})$, we can say that

the tensor product linearizes what was bilinear (or multilinear).

There is no reason to restrict oneself to the tensor product of only two factors. One can equally define the tensor product $V_1\otimes\cdots\otimes V_k$, and obtain a vector space of dimension $\dim V_1\times\cdots\times\dim V_k$. Note that we do not need to use brackets, since the tensor product is associative: $(V_1\otimes V_2)\otimes V_3 = V_1\otimes(V_2\otimes V_3)$.

We have

$$\mathcal{T}^p_q(V) = \underbrace{V\otimes\cdots\otimes V}_{p}\otimes\underbrace{V^*\otimes\cdots\otimes V^*}_{q},$$

since both spaces have the same basis. An element T of $\mathcal{T}^p_q(V)$ was first regarded according to Definition 4.8 as a multilinear map

$$T:\underbrace{V^*\times\ldots V^*}_{p}\times\underbrace{V\times\cdots\times V}_{q}\longrightarrow\mathbb{R}\,.$$

Now, with respect to bases $\mathcal{B}=\{b_1,\ldots,b_n\}$ of V and $\mathcal{B}^*=\{\beta^1,\ldots,\beta^n\}$ of V^*, the components of T are

$$T^{i_1,\ldots,i_p}_{j_1,\ldots,j_q} := T(\beta^{i_1},\ldots,\beta^{i_p},b_{j_1},\ldots,b_{j_q})\,,$$

hence we may view T as

$$T^{i_1,\ldots,i_p}_{j_1,\ldots,j_q}b_{i_1}\otimes\ldots\otimes b_{i_p}\otimes\beta^{j_1}\otimes\ldots\otimes\beta^{j_q}\in\underbrace{V\otimes\cdots\otimes V}_{p}\otimes\underbrace{V^*\otimes\cdots\otimes V^*}_{q}\,.$$

In particular, the pth **tensor power** of a vector space is

$$V^{\otimes p} := \underbrace{V\otimes\cdots\otimes V}_{p} = \mathcal{T}^p_0(V)\,.$$

Therefore, we may write

$$\boxed{\mathrm{Bil}(V^*\times V^*,\mathbb{R}) = V\otimes V = V^{\otimes 2}},\quad \boxed{\mathrm{Bil}(V\times V,\mathbb{R}) = V^*\otimes V^* = (V^*)^{\otimes 2}}$$

and

$$\boxed{\mathcal{T}^p_q(V) = V^{\otimes p}\otimes(V^*)^{\otimes q}}.$$

CHAPTER 5

Applications

Mathematically-speaking, a tensor is a real-valued function of some number of vectors and some number of covectors (a.k.a. linear forms), which is linear in each of its arguments. On the other hand, tensors are most useful in connection with concrete physical applications. In this chapter, we borrow notions and computations from Physics and Calculus to discuss important classical tensors.

5.1. Inertia Tensor

5.1.1. Physical Preliminaries.

We consider a rigid body M fixed at a point O and rotating about an axis through O with **angular velocity** ω. Denoting the time variable t and an angle variable θ around the axis of rotation, the angular velocity will be viewed as a vector[18] with magnitude

$$\|\omega\| = \left\| \frac{d\theta}{dt} \right\| ,$$

with direction given by the axis of rotation and with orientation given by the right-hand rule. The **position vector** of a point P in the body M relative to the origin O is

$$\mathbf{x} := \overrightarrow{OP}$$

while the **linear velocity** of that point P is

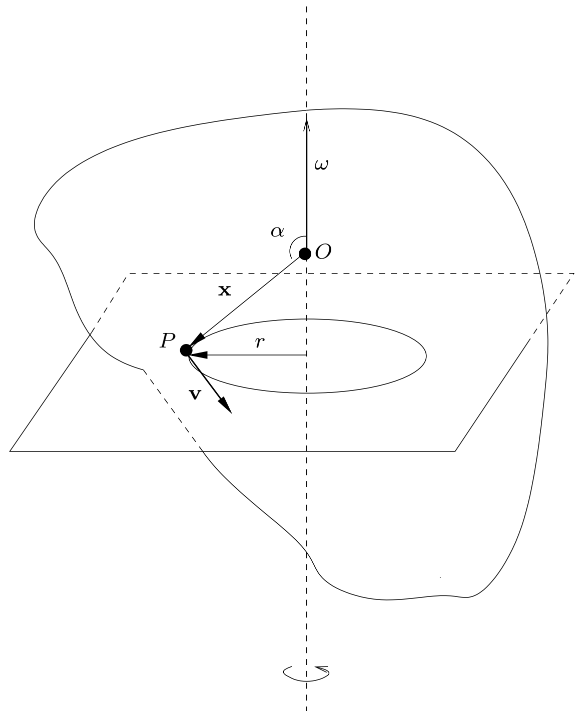
$$\mathbf{v} := \omega \times \mathbf{x} .$$

The linear velocity $\mathbf{v}$ has, hence, magnitude

$$\|\mathbf{v}\| = \underbrace{\|\omega\|}_{\|\frac{d\theta}{dt}\|} \underbrace{\|\mathbf{x}\| \sin\alpha}_{=:r} ,$$

where α is the angle between ω and $\mathbf{x}$, and has direction tangent at P to the circle of radius r perpendicular to the axis of rotation.

[18]Warning: The angular velocity is actually only what physicists call a *pseudovector* because it does not follow the usual contravariance of a vector in case of orientation flip. Luckily, this issue does not affect the inertia tensor, since the sign flip cancels out thanks to squaring.



The **kinetic energy** of an infinitesimal region dM of M around P is

$$dE = \frac{1}{2}\|\mathbf{v}\|^2 dm\,,$$

where $\|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v}$ is the square of the norm of the linear velocity and dm is the mass of dM. The **total kinetic energy** of M is

$$E = \frac{1}{2}\int_M \|\mathbf{v}\|^2 dm = \frac{1}{2}\int_M \|\omega\times\mathbf{x}\|^2 dm\,.$$

Actually, depending on the type of rigid body, we might take here a sum instead of integral, or some other type of integral (line integral, surface integral, etc) such as:

(1) If M is a solid in 3-dimensional space, then

$$E = \frac{1}{2}\iiint_M \|\omega\times\mathbf{x}_P\|^2 \rho_P\, dx^1dx^2dx^3\,,$$

where the norm squared $\|\omega\times\mathbf{x}_P\|^2$ and the density ρ_P are functions of the point P with coordinates (x^1,x^2,x^3).

(2) If M is a flat sheet, then

$$E = \frac{1}{2}\iint_M \|\omega\times\mathbf{x}_P\|^2 \rho_P\, dx^1dx^2\,,$$

where the integrand only depends on two cartesian coordinates.

(3) If M is a (curvy) surface in 3-dimensional space, then

$$E = \frac{1}{2} \iint_M \|\omega \times \mathbf{x}_P\|^2 \rho_P \, d\sigma \,,$$

where $d\sigma$ is the infinitesimal element of the surface for a surface integral.

(4) If M is a wire in 3-dimensional space, then

$$E = \frac{1}{2} \int_M \|\omega \times \mathbf{x}_P\|^2 \rho_P \, ds \,,$$

where ds is the infinitesimal element of length for a line integral.

(5) If M is a finite set of N point masses m_i with rigid relative positions, then

$$E = \frac{1}{2} \sum_{i=1}^{N} \|\omega \times \mathbf{x_i}\|^2 \, m_i \,.$$

We will keep writing our formulas for the first case (with a volume integral); these should be adjusted for situations of the other types.

In any case, we need to work out the quantity

$$\|\omega \times \mathbf{x}\|^2$$

for vectors ω and $\mathbf{x}$ in 3-dimensional space.

To this purpose, we use the **Lagrange identity**[19], according to which

$$(a \times b) \cdot (c \times d) = \det \begin{bmatrix} a \cdot c & a \cdot d \\ b \cdot c & b \cdot d \end{bmatrix} \,. \tag{5.1}$$

Applying (5.1) with $a = c = \omega$ and $b = d = \mathbf{x}$, we obtain

$$\|\omega \times \mathbf{x}\|^2 = (\omega \times \mathbf{x}) \cdot (\omega \times \mathbf{x}) = \det \begin{bmatrix} \omega \cdot \omega & \omega \cdot \mathbf{x} \\ \mathbf{x} \cdot \omega & \mathbf{x} \cdot \mathbf{x} \end{bmatrix} = \|\omega\|^2 \|\mathbf{x}\|^2 - \|\omega \cdot \mathbf{x}\|^2 \,.$$

Let now $\mathcal{B} = \{e_1, e_2, e_3\}$ be an orthonormal[20] basis of $\mathbb{R}^3$, so that

$$\omega = \omega^i e_i \quad \text{and} \quad \mathbf{x} = x^i e_i \,.$$

Then

$$\begin{aligned} \|\omega\|^2 &= \omega \cdot \omega = \delta_{ij} \omega^i \omega^j = \omega^1 \omega^1 + \omega^2 \omega^2 + \omega^3 \omega^3 \\ \|\mathbf{x}\|^2 &= \mathbf{x} \cdot \mathbf{x} = \delta_{k\ell} x^k x^\ell = x^1 x^1 + x^2 x^2 + x^3 x^3 \\ \omega \cdot \mathbf{x} &= \delta_{ik} \omega^i x^k = \omega^1 x^1 + \omega^2 x^2 + \omega^3 x^3 \end{aligned}$$

[19] The Lagrange identity can be patiently proven in coordinates.

[20] We could use any basis of $\mathbb{R}^3$. Then, instead of the δ_{ij}, the formulas would have involved the components of the metric tensor g_{ij}. However, computations with orthonormal bases are simpler; in particular, the inverse of an orthonormal basis change L is simply ${}^{t}L$. Moreover, the inertia tensor is symmetric, hence admits an orthonormal eigenbasis.

so that

$$\begin{aligned}\|\omega \times \mathbf{x}\|^2 &= \|\omega\|^2\|\mathbf{x}\|^2 - \|\omega \cdot \mathbf{x}\|^2 \\ &= (\delta_{ij}\omega^i\omega^j)(\delta_{k\ell}x^k x^\ell) - (\delta_{ik}\omega^i x^k)(\delta_{jl}\omega^j x^\ell) \\ &= (\delta_{ij}\delta_{k\ell} - \delta_{ik}\delta_{j\ell})\omega^i\omega^j x^k x^\ell \,.\end{aligned}$$

Therefore, the total kinetic energy is

$$\boxed{E = \frac{1}{2}(\delta_{ij}\delta_{k\ell} - \delta_{ik}\delta_{j\ell})\omega^i\omega^j \iiint_M x^k x^\ell dm}$$

and it depends only on $\omega^1, \omega^2, \omega^3$ (since we have integrated over the x^1, x^2, x^3).

5.1.2. Moments of Inertia.

Definition 5.1. The **inertia tensor** is the covariant 2-tensor whose components with respect to an orthonormal basis $\mathcal{B}$ are

$$\boxed{I_{ij} = (\delta_{ij}\delta_{k\ell} - \delta_{ik}\delta_{j\ell}) \iiint_M x^k x^\ell dm}\,.$$

Then the kinetic energy of the rotating rigid body is

$$\boxed{E = \frac{1}{2} I_{ij}\omega^i\omega^j}$$

which in matrix notation amounts to

$$E = \frac{1}{2}\omega \cdot I\omega = \frac{1}{2}{}^{\mathrm{t}}\omega I\omega \,.$$

Remark 5.2. If, instead of an orthonormal basis, we had used any basis of $\mathbb{R}^3$, we would have gotten

$$I_{ij} = (g_{ij}g_{k\ell} - g_{ik}g_{j\ell}) \iiint_M x^k x^\ell dm \,.$$

where g_{ij} are the components of the metric tensor. This formula also makes apparent the covariance and the symmetry of I inherited from the metric: $I_{ij} = I_{ji}$ for all i and j.

We will see that the inertia tensor is a convenient way to encode all moments of inertia of an object in one quantity and we return now to the case of an orthonormal basis. The first component of the inertia tensor is

$$I_{11} = (\underbrace{\delta_{11}\delta_{k\ell}}_{\substack{=0 \\ \text{unless} \\ k=\ell}} - \underbrace{\delta_{1k}\delta_{1\ell}}_{\substack{=0 \\ \text{unless} \\ k=\ell=1}}) \iiint_M x^k x^\ell dm \,.$$

If $k = \ell = 1$, then $\delta_{11}\delta_{11} - \delta_{11}\delta_{11} = 0$, so that the non-vanishing terms have $k = \ell \neq 1$. In this way, one can check that

$$\begin{aligned}
I_{11} &= \iiint_M (x^2x^2 + x^3x^3)\ dm \\
I_{22} &= \iiint_M (x^1x^1 + x^3x^3)\ dm \\
I_{33} &= \iiint_M (x^1x^1 + x^2x^2)\ dm \\
I_{23} = I_{32} &= -\iiint_M x^2x^3\ dm \\
I_{31} = I_{13} &= -\iiint_M x^1x^3\ dm \\
I_{12} = I_{21} &= -\iiint_M x^1x^2\ dm\,,
\end{aligned}$$

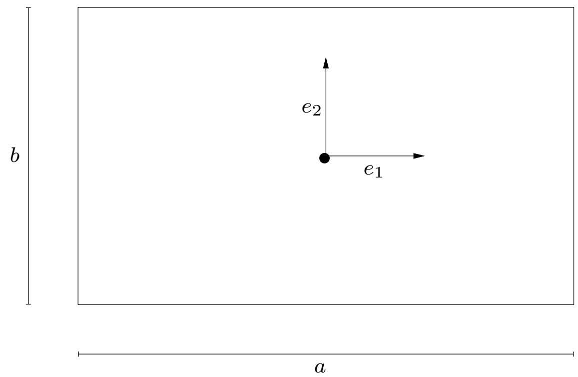
so that with respect to an orthonormal basis $\mathcal{B}$, the inertia tensor is represented by the symmetric matrix

$$I = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix} .$$

The diagonal components I_{11}, I_{22}, I_{33} are the **moments of inertia** of the rigid body M with respect to the coordinate axes Ox_1, Ox_2, Ox_3, respectively. The off-diagonal components I_{12}, I_{23}, I_{31} are the **polar moments of inertia** or the **products of inertia** of the rigid body M.

Example 5.3. We want to find the inertia tensor of a homogeneous rectangular plate with sides a and b and total mass m, assuming that the rotation preserves the center of mass O. We choose a coordinate system (corresponding to an orthonormal basis) with origin at the center of mass O, with x-axis parallel to the side of length a, y-axis parallel to the side of length b, z-axis perpendicular to the plate, and adjust our previous formulas to double integrals. Since the plate is assumed to be homogeneous, it has a constant *mass density* equal to

$$\rho = \frac{\text{total mass}}{\text{area}} = \frac{m}{ab}\,.$$



Then

$$\underbrace{I_{11}}_{I_{xx}} = \int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\frac{b}{2}}^{\frac{b}{2}} (y^2 + \underbrace{z^2}_{=0}) \underbrace{\rho}_{\frac{m}{ab}} \, dy \, dx$$

$$= \frac{m}{ab} a \int_{-\frac{b}{2}}^{\frac{b}{2}} y^2 \, dy$$

$$= \frac{m}{b} \left[\frac{y^3}{3} \right]_{-\frac{b}{2}}^{\frac{b}{2}} = \frac{m}{12} b^2 .$$

Similarly,

$$\underbrace{I_{22}}_{I_{yy}} = \frac{m}{12} a^2 ,$$

and

$$\underbrace{I_{33}}_{I_{zz}} = \int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\frac{b}{2}}^{\frac{b}{2}} (x^2 + y^2) \rho \, dy \, dx = \frac{m}{12}(a^2 + b^2)$$

turns out to be just the sum of I_{11} and I_{22}.

Furthermore,

$$I_{23} = I_{32} = -\int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\frac{b}{2}}^{\frac{b}{2}} y \underbrace{z}_{=0} \rho \, dy \, dx = 0 ,$$

and, similarly, $I_{31} = I_{13} = 0$. Finally, we have

$$I_{21} = I_{21} = -\int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\frac{b}{2}}^{\frac{b}{2}} xy \, \rho \, dy \, dx = -\frac{m}{ab} \underbrace{\underbrace{\left(\int_{-\frac{a}{2}}^{\frac{a}{2}} x \, dx \right)}_{=0} \underbrace{\left(\int_{-\frac{b}{2}}^{\frac{b}{2}} y \, dy \right)}_{=0}}_{\text{because the integral of an odd function on a symmetric interval is } 0} .$$

We conclude that the inertia tensor is given by the matrix

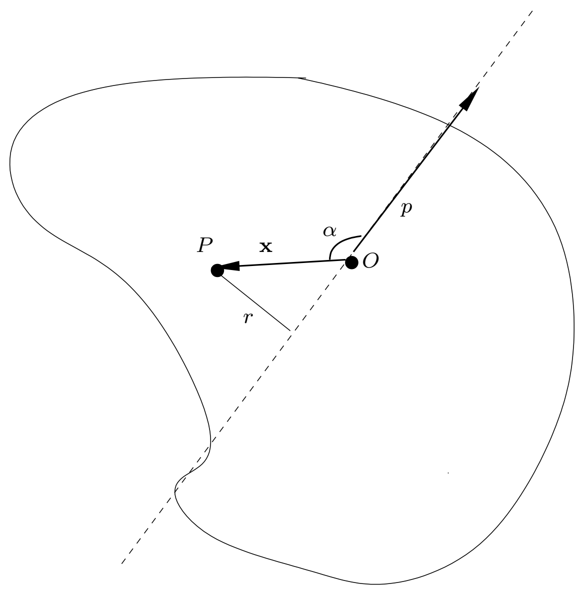
$$\frac{m}{12}\begin{pmatrix} b^2 & 0 & 0 \\ 0 & a^2 & 0 \\ 0 & 0 & a^2+b^2 \end{pmatrix} .$$

□

EXERCISE 5.4. Compute the inertia tensor of the same plate, but now with center of rotation O coinciding with a vertex of the rectangular plate.

5.1.3. Moment of Inertia About any Axis.

We compute the moment of inertia of the body M about an axis through the point O and defined by the unit vector p.



The **moment of inertia** of an infinitesimal region of M around P is

$$dI \;=\; \underbrace{r^2}_{\substack{r \text{ is the distance} \\ \text{from } P \text{ to the axis}}} \quad \underbrace{dm}_{\substack{\text{infinitesimal} \\ \text{mass}}} \;=\; \|p \times \mathbf{x}\|^2 \, dm \,,$$

where the last equality follows from the fact that $\|p \times x\| = \|p\| \, \|x\| \sin\alpha = r$, since p is a unit vector. Hence, the **total moment of inertia** of M with respect to the axis given by p is

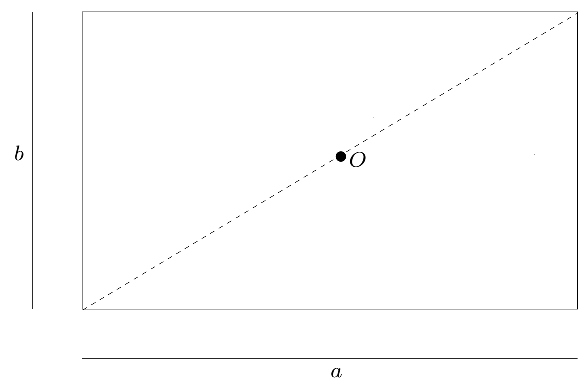
$$I_p := \iiint_M \|p \times \mathbf{x}\|^2 \, dm \;\geq\; 0 \,.$$

This is very similar to the total kinetic energy E: just replace ω by p and omit the factor $\frac{1}{2}$. By the earlier computations, we conclude that

$$I = I_{ij} p^i p^j \,,$$

where I_{ij} is the inertia tensor. This formula shows that *the total moment of inertia of the rigid body M with respect to an arbitrary axis passing through the point O is determined only by the inertia tensor of the rigid body.*

EXAMPLE 5.5. For the rectangular plate in Example 5.3, we now want to compute the moment of inertia with respect to the diagonal of the plate.



We choose the unit vector $p = \frac{1}{\sqrt{a^2+b^2}}(ae_1 + be_2)$ (the other possibility is the negative of this vector, yielding the same result), so that

$$p^1 = \frac{a}{\sqrt{a^2+b^2}}\,, \quad p^2 = \frac{b}{\sqrt{a^2+b^2}}\,, \quad p^3 = 0$$

and use the matrix for I found in Example 5.3. The moment of inertia is

$$\begin{aligned} I_p =& I_{ij}p^ip^j \\ =& \begin{pmatrix} \frac{a}{\sqrt{a^2+b^2}} & \frac{b}{\sqrt{a^2+b^2}} & 0 \end{pmatrix} \begin{pmatrix} \frac{m}{12}b^2 & 0 & 0 \\ 0 & \frac{m}{12}a^2 & 0 \\ 0 & 0 & \frac{m}{12}(a^2+b^2) \end{pmatrix} \begin{pmatrix} \frac{a}{\sqrt{a^2+b^2}} \\ \frac{b}{\sqrt{a^2+b^2}} \\ 0 \end{pmatrix} \\ =& \frac{m}{6}\frac{a^2b^2}{a^2+b^2}\,. \end{aligned}$$

□

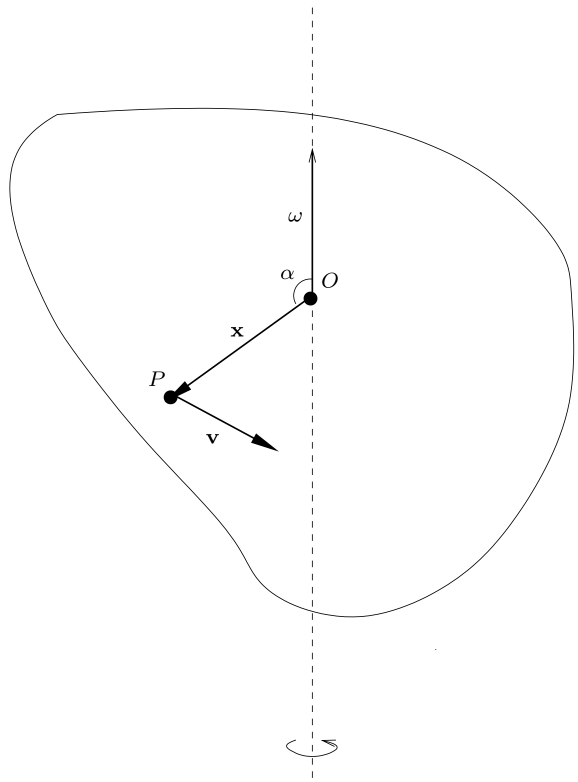
EXERCISE 5.6. Double-check the above result for the moment of inertia of the rectangular plate in Example 5.3 with respect to the diagonal of the plate, now by using the inertia tensor computed in Exercise 5.4 (with center of rotation O in a vertex belonging also to that diagonal).

EXERCISE 5.7. Compute the moment of inertia of the rectangular plate in Example 5.3 with respect to an axis perpendicular to the plate and passing through its center of mass.

EXERCISE 5.8. Compute the moment of inertia of the rectangular plate in Example 5.3 with respect to an axis perpendicular to the plate and passing through one vertex.

5.1.4. Angular Momentum.

Let M be a body rotating with angular velocity ω about an axis through the point O. Let $\mathbf{x} = \overrightarrow{OP}$ be the position vector of a point P and $\mathbf{v} = \omega \times \mathbf{x}$ the linear velocity of P.



Then the **angular momentum** of an infinitesimal region of M around P is

$$dL = (\mathbf{x} \times \mathbf{v})\, dm\,,$$

so that the **total angular momentum**[21] of M is

$$L = \iiint_M (\mathbf{x} \times (\omega \times \mathbf{x}))\, dm\,.$$

[21]Just like the angular velocity, the angular momentum is not an honest vector, but only a *pseudovector*, since there is an issue with orientation. In this subsection, we should thus assume that we work with an *oriented* orthonormal basis $\{e_1, e_2, e_3\}$ of $\mathbb{R}^3$, so that $e_1 \times e_2 = e_3$ (and not $-e_3$). This amounts to assuming that the change of basis matrix L from the standard basis has $\det L = 1$ (and not -1).

We need to work out $\mathbf{x} \times (\omega \times \mathbf{x})$ for vectors $\mathbf{x}$ and ω in 3-dimensional space. We apply the following identity[22] for the triple vector product:

$$\mathbf{x} \times (\omega \times \mathbf{x}) = (\mathbf{x} \cdot \mathbf{x})\omega - (\omega \cdot \mathbf{x})\mathbf{x}\,. \tag{5.2}$$

Let $\mathcal{B} = \{e_1, e_2, e_3\}$ be an orthonormal basis of $\mathbb{R}^3$. Then, replacing the following equalities

$$\omega = \omega^i e_i = \delta^i_j \omega^j e_i \tag{5.3}$$

$$\mathbf{x} = x^i e_i = \delta^i_k x^k e_i \tag{5.4}$$

$$\mathbf{x} \cdot \mathbf{x} = \delta_{k\ell} x^k x^\ell \tag{5.5}$$

$$\omega \cdot \mathbf{x} = \delta_{j\ell} \omega^j x^\ell \tag{5.6}$$

into (5.2), we obtain

$$\mathbf{x} \times (\omega \times \mathbf{x}) = (\underbrace{\delta_{k\ell} x^k x^\ell}_{(5.5)})\, \underbrace{\delta^i_j \omega^j e_i}_{(5.3)} - (\underbrace{\delta_{j\ell} \omega^j x^\ell}_{(5.6)})\, \underbrace{\delta^i_k x^k e_i}_{(5.4)} = (\delta^i_j \delta_{k\ell} - \delta^i_k \delta_{j\ell}) \omega^j x^k x^\ell e_i\,.$$

Therefore, the total angular momentum is

$$L = L^i e_i\,,$$

where the components L^i are

$$\boxed{L^i = (\delta^i_j \delta_{k\ell} - \delta^i_k \delta_{j\ell})\, \omega^j \iiint_M x^k x^\ell \, dm}\,.$$

Since we have always

$$\delta^i_j = \delta_{ij} = \delta^{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

if we violate index balance for one instance, the above expression for L^i can be written in terms of the inertia tensor I_{ij} as

$$L^i = I_{ij} \omega^j\,,$$

which corresponds to the matrix form $L = I\omega$. We see that the angular momentum L is proportional (or parallel) to the angular velocity ω only when ω is an eigenvector of the inertia tensor I.

Example 5.9. Suppose the rectangular plate in the previous examples is rotating about an axis through the center of mass O with angular velocity

$$\omega = e_1 + 2e_2 + 3e_3\,, \quad \text{or} \quad \begin{pmatrix} \omega^1 \\ \omega^2 \\ \omega^3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

We want to compute its angular momentum.

[22] To prove this vector equality use coordinates, consider only the case in which ω is a standard basis vector and then use the linearity in ω.

The inertia tensor is given by the matrix I_{ij} found in Example 5.3:

$$\begin{pmatrix} \frac{m}{12}b^2 & 0 & 0 \\ 0 & \frac{m}{12}a^2 & 0 \\ 0 & 0 & \frac{m}{12}(a^2+b^2) \end{pmatrix} .$$

The total angular momentum has components given by

$$\begin{pmatrix} L^1 \\ L^2 \\ L^3 \end{pmatrix} = \begin{pmatrix} \frac{m}{12}b^2 & 0 & 0 \\ 0 & \frac{m}{12}a^2 & 0 \\ 0 & 0 & \frac{m}{12}(a^2+b^2) \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{m}{12}b^2 \\ \frac{m}{6}a^2 \\ \frac{m}{4}(a^2+b^2) \end{pmatrix} ,$$

so that

$$L = \frac{m}{12}b^2 e_1 + \frac{m}{6}a^2 e_2 + \frac{m}{4}(a^2+b^2)e_3 .$$

□

5.1.5. Principal Moments of Inertia.

Observe that the inertia tensor of a rigid body M is *symmetric* and recall the *spectral theorem* (Theorem 3.9). Then we know that an orthonormal eigenbasis $\{\tilde{e}_1, \tilde{e}_2, \tilde{e}_3\}$ exists for the inertia tensor. Let I_1, I_2, I_3 be the corresponding eigenvalues. The matrix representing the inertia tensor with respect to this eigenbasis is

$$\begin{pmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{pmatrix} .$$

The orthonormal eigenbasis gives a preferred coordinate system in which to formulate a problem pertaining to rotation of this body. The axes of the eigenvectors are called the **principal axes of inertia** of the rigid body M. The eigenvalues I_i are called the **principal moments of inertia**.

For instance, if a homogeneous body is symmetric with respect to the xy-plane, then the polar moments of inertia $I_{23} = I_{32}$ and $I_{13} = I_{31}$ vanish, thus the z-axis is necessarily a principal axis (because of the block-form of I).

The *principal moments of inertia* are the moments of inertia with respect to the *principal axes of inertia*, hence they are non-negative

$$I_1, , I_2, , I_3 \geq 0 .$$

A rigid body is called

(1) an **asymmetric top** if $I_1 \neq I_2 \neq I_3 \neq I_1$;
(2) a **symmetric top** if exactly two eigenvalues are equal, say $I_1 = I_2 \neq I_3$: any axis passing through the plane determined by $\tilde{e}_1$ and $\tilde{e}_2$ is then a principal axis of inertia;
(3) a **spherical top** if $I_1 = I_2 = I_3$: any axis passing through O is a principal axis of inertia.

With respect to the eigenbasis $\{\tilde{e}_1, \tilde{e}_2, \tilde{e}_3\}$ the *kinetic energy* is

$$E = \frac{1}{2}(I_1(\tilde{\omega}^1)^2 + I_2(\tilde{\omega}^2)^2 + I_3(\tilde{\omega}^3)^2),$$

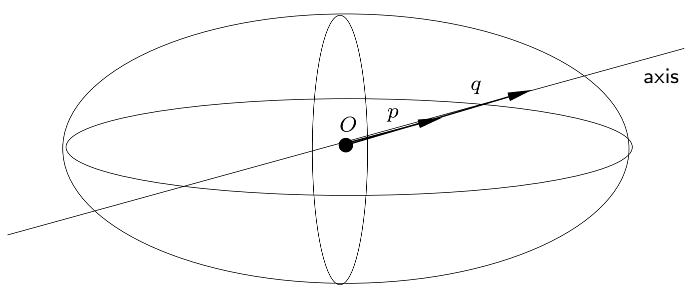
where $\omega = \tilde{\omega}^i \tilde{e}_i$, with $\tilde{\omega}^i$ the components of the angular velocity with respect to the basis $\{\tilde{e}_1, \tilde{e}_2, \tilde{e}_3\}$. In particular, we see that the kinetic energy can be conserved, even if the angular velocity ω changes, as long as the above combination of squares is preserved. This is related to the phenomenon of *precession*.

The surface determined by the equation (with respect to the coordinates x, y, z)

$$I_1 x^2 + I_2 y^2 + I_3 z^2 = 1 \tag{5.7}$$

is called the **ellipsoid of inertia**. The symmetry axes of the ellipsoid coincide with the principal axes of inertia. Note that for a spherical top, the ellipsoid of inertia is actually a sphere.

The ellipsoid of inertia gives the moment of inertia with respect to any axis as follows: Consider an axis given by the unit vector p and let $q = cp$ be a vector of intersection of the axis with the ellipsoid of inertia, where c is $(\pm)$ the distance to O of the intersection of the axis with the ellipsoid of inertia.



The moment of inertia with respect to this axis is

$$I = I_{ij} p^i p^j = \frac{1}{c^2} I_{ij} q^i q^j = \frac{1}{c^2},$$

where the last equality follows from the fact that, since q is on the ellipsoid, then $I_{ij} q^i q^j = 1$ by equation (5.7).

Example 5.10. The principal axes of inertia for the rectangular plate in Example 5.3 are the axes parallel to the sides and the axis perpendicular to the plate. The corresponding principal moments of inertia are

$$I_{11} = \frac{m}{12} b^2, \qquad I_{22} = \frac{m}{12} a^2 \quad \text{and} \quad I_{33} = \frac{m}{12}(a^2 + b^2).$$

If $a = b$, that is, if the rectangle is a square, we have a *symmetric top*. □

5.2. Stress Tensor (Spannung)

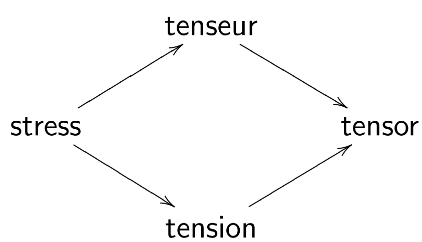
5.2.1. Physical Preliminaries.

Let us consider a rigid body M acted upon by external forces but in *static equilibrium*, and let us consider an infinitesimal region dM around a point P. There are two types of external forces:

(1) The **body forces**, that is forces whose magnitude is proportional to the volume/mass of the region. For instance, *gravity*, *attractive force* or the *centrifugal force*.
(2) The **surface forces**, that is forces exerted on the surface of the element by the material surrounding it. These are forces whose magnitude is proportional to the area of the region in consideration.

The *surface force per unit area* is called the **stress**. We will concentrate on **homogeneous stress**, that is stress that does not depend on the location of the element in the body, but depends only on the orientation of the surface given by its tangent plane. Moreover, we assume that the body in consideration is in *static equilibrium*.

REMARK 5.11. It was the concept of *stress* in mechanics that originally led to the invention of tensors



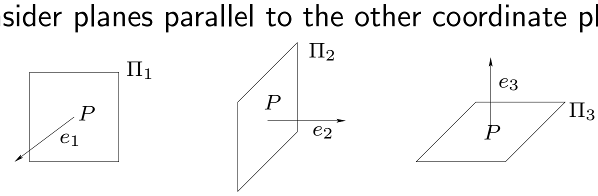
Choose an orthonormal basis $\{e_1, e_2, e_3\}$ and the plane Π through P parallel to the e_2e_3 coordinate plane. The vector e_1 is normal to this plane. Let ΔA_1 be the area of the slice of the infinitesimal region around P cut by the plane and let ΔF be the force acting on that slice. We write ΔF in terms of its components

$$\Delta F = \Delta F^1 e_1 + \Delta F^2 e_e + \Delta F^3 e_3$$

and, since the stress is the surface force per unit area, we define

$$\sigma^{1j} := \lim_{\Delta A_1 \to 0} \frac{\Delta F^j}{\Delta A_1}, \qquad \text{for } j = 1, 2, 3\,.$$

Similarly, we can consider planes parallel to the other coordinate planes

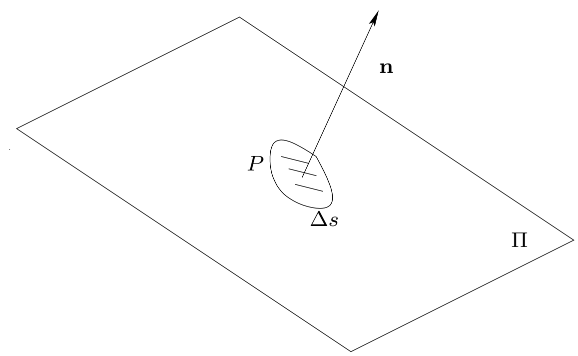


and define

$$\boxed{\sigma^{ij} := \lim_{\Delta A_i \to 0} \frac{\Delta F^j}{\Delta A_i}}.$$

It turns out that the resulting nine numbers σ^{ij} are the components of a contravariant 2-tensor called the **stress tensor**, as we will see in §5.2.5. The stress tensor encodes the mechanical stresses on an object.

We now compute the stress across *other* slices through P, that is, across other planes with other normal vectors. Let Π be a plane passing through P, $\mathbf{n}$ a unit vector through P perpendicular to the plane Π, $\Delta s = \Pi \cap dM$ the area of a small element of the plane Π containing P and ΔF the force acting on that element.

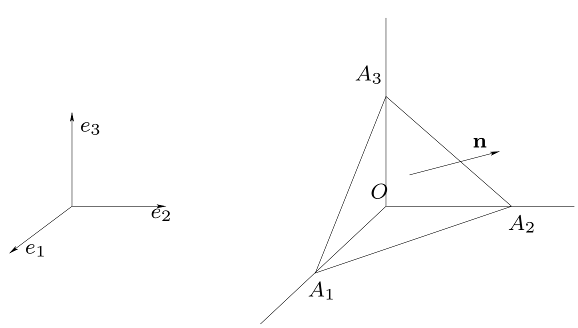


CLAIM 5.12. The stress at P across the surface perpendicular to $\mathbf{n}$ is

$$\sigma(\mathbf{n}) := \lim_{\Delta s \to 0} \frac{\Delta F}{\Delta s} = \sigma^{ij}(\mathbf{n} \cdot e_i)e_j \,.$$

It follows from the claim that the stress σ is a vector-valued function that depends linearly on the normal $\mathbf{n}$ to the surface element.

PROOF. Consider the tetrahedron $OA_1A_2A_3$ bound by the *triangular slice* on the plane Π having area Δs and three triangles on planes parallel to the coordinate planes



Consider all external forces acting on this tetrahedron, which we regard as a volume element of the rigid body:

(1) *Body forces* amounting to $f \cdot \Delta v$, where f is the force per unit of volume and Δv is the volume of the tetrahedron. We actually do not know these forces, but we will see later that these are not relevant.
(2) *Surface forces* amounting to the sum of the forces on each of the four sides of the tetrahedron.

We want to assess each of the four surface contributions due to the surface forces. If Δs is the area of the slice on the plane Π, the contribution of that slice is, by definition of stress equal to

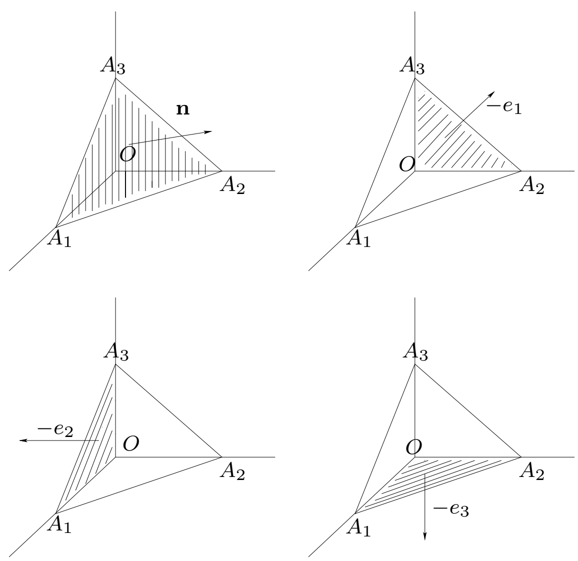
$$\sigma(\mathbf{n})\Delta s\,.$$

If Δs_1 is the area of the slice on the plane with normal $-e_1$, the contribution of that slice is

$$-\sigma^{1j}e_j\Delta s_1\,,$$

and, similarly, the contributions of the other two slices are

$$-\sigma^{2j}e_j\Delta s_2 \qquad \text{and} \qquad -\sigma^{3j}e_j\Delta s_3\,.$$



Note that the minus sign comes from the fact that we use everywhere outside pointing normals.

So the total surface force is

$$\sigma(\mathbf{n})\Delta s - \sigma^{1j}e_j\Delta s_1 - \sigma^{2j}e_j\Delta s_2 - \sigma^{3j}e_j\Delta s_3\,.$$

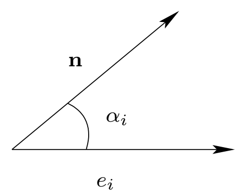
Since there is static equilibrium, the sum of all (body and surface) forces must be zero

$$f\Delta v + \sigma(\mathbf{n})\Delta s - \sigma^{ij}e_j\Delta s_i = 0\,.$$

The term $f\Delta v$ can be neglected when Δs is small, as it contains terms of higher order (in fact, $\Delta v \to 0$ faster than $\Delta s \to 0$). We conclude that

$$\sigma(\mathbf{n})\Delta s = \sigma^{ij}e_j\Delta s_i\,.$$

It remains to relate Δs to $\Delta s_1, \Delta s_2, \Delta s_3$. The side with area Δs_i is the orthogonal projection of the side with area Δs onto the plane with normal e_i. The scaling factor for the area under projection is $\cos\alpha_i$, where α_i is the convex angle between the plane normal vectors



$$\frac{\Delta s_i}{\Delta s} = \cos\alpha_i = \cos\alpha_i \|\mathbf{n}\| \, \|e_i\| = \mathbf{n} \cdot e_i \, .$$

Therefore ,

$$\sigma(\mathbf{n})\Delta s = \sigma^{ij} e_j (\mathbf{n} \cdot e_i) \Delta s$$

or, equivalently,

$$\sigma(\mathbf{n}) = \sigma^{ij} (\mathbf{n} \cdot e_i) e_j \, .$$

Note that, in the above formula, the quantities $\mathbf{n} \cdot e_i$ are the coordinates of $\mathbf{n}$ with respect to the orthonormal basis $\{e_1, e_2, e_3\}$, namely

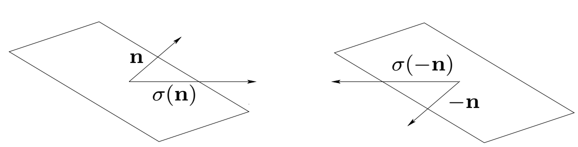
$$\mathbf{n} = (\mathbf{n} \cdot e_1)e_1 + (\mathbf{n} \cdot e_2)e_2 + (\mathbf{n} \cdot e_3)e_3 = n^1 e_1 + n^2 e_2 + n^3 e_3 \, .$$

□

Remark 5.13. For homogeneous stress, the stress tensor σ^{ij} does not depend on the point P. However, when we flip the orientation of the normal to the plane, the stress tensor changes sign. In other words, if $\sigma(\mathbf{n})$ is the stress across a surface with normal $\mathbf{n}$, then

$$\sigma(-\mathbf{n}) = -\sigma(\mathbf{n}) \, .$$

The stress considers orientation as if the forces on each side of the surface have to balance each other in static equilibrium.

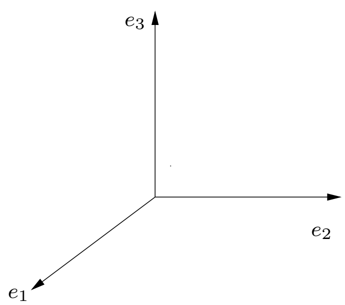


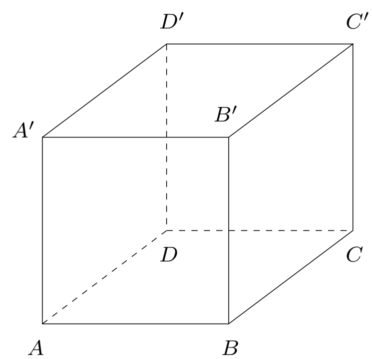
□

5.2.2. Principal Stresses.

CLAIM 5.14. The stress tensor is a *symmetric tensor*, that is $\sigma^{ij} = \sigma^{ji}$.

PROOF. Consider an infinitesimal cube of side $\Delta\ell$ surrounding P and with faces parallel to the coordinate planes.





The force acting on each of the six faces of the cube are:

- $\sigma^{1j}\Delta A_1 e_j$ and $-\sigma^{1j}\Delta A_1 e_j$, respectively for the front and the back faces, $ABB'A'$ and $DCC'D'$;
- $\sigma^{2j}\Delta A_2 e_j$ and $-\sigma^{2j}\Delta A_2 e_j$, respectively for the right and the left faces $BCC'B'$ and $ADD'A'$;
- $\sigma^{3j}\Delta A_3 e_j$ and $-\sigma^{3j}\Delta A_3 e_j$, respectively for the top and the bottom faces $ABCD$ and $A'B'C'D'$,

where $\Delta A_1 = \Delta A_2 = \Delta A_3 = \Delta s = (\Delta\ell)^2$ is the common face area. We compute now the *torque* μ, assuming the forces are applied at the center of each face, whose distance to the center point P is $\frac{1}{2}\Delta\ell$. Recall that the **torque** is the tendency of a force to twist or rotate an object. It is given by the cross product of the distance vector and the force vector.

$$
\begin{aligned}
\mu =& \tfrac{\Delta\ell}{2} e_1 \times \sigma^{1j}\Delta s\, e_j + \left(-\tfrac{\Delta\ell}{2} e_1\right) \times (-\sigma^{1j}\Delta s\, e_j\,) \\
&+ \tfrac{\Delta\ell}{2} e_2 \times \sigma^{2j}\Delta s\, e_j + \left(-\tfrac{\Delta\ell}{2} e_2\right) \times (-\sigma^{2j}\Delta s\, e_j\,) \\
&+ \tfrac{\Delta\ell}{2} e_2 \times \sigma^{3j}\Delta s\, e_j + \left(-\tfrac{\Delta\ell}{2} e_3\right) \times (-\sigma^{3j}\Delta s\, e_j\,) \\
=& \Delta\ell\Delta s\,(e_i \times \sigma^{ij} e_j) \\
=& \Delta\ell\Delta s\big((\sigma^{23} - \sigma^{32})e_1 + (\sigma^{31} - \sigma^{13})e_2 + (\sigma^{12} - \sigma^{21})e_3\big)\,.
\end{aligned}
$$

Since the equilibrium is static, then $\mu = 0$, so that $\sigma^{ij} = \sigma^{ji}$. □

We can hence write

$$\sigma = \begin{pmatrix} \sigma^{11} & \sigma^{12} & \sigma^{13} \\ \sigma^{12} & \sigma^{22} & \sigma^{23} \\ \sigma^{13} & \sigma^{23} & \sigma^{33} \end{pmatrix},$$

where the diagonal entries σ^{11}, σ^{22} and σ^{33} are the **normal components**, that is the components of the forces perpendicular to the coordinate planes and the remaining entries σ^{12}, σ^{13} and σ^{23} are the **shear components**, that is the components of the forces parallel to the coordinate planes.

Since the stress tensor is symmetric, again by the *spectral theorem* (Theorem 3.9), it can be orthogonally diagonalized, that is

$$\sigma = \begin{pmatrix} \sigma^{1} & 0 & 0 \\ 0 & \sigma^{2} & 0 \\ 0 & 0 & \sigma^{3} \end{pmatrix},$$

where now σ^1, σ^2 and σ^3 are the **principal stresses**, that is the *eigenvalues of* σ. The eigenspaces of σ are the **principal directions** and the shear components disappear for the **principal planes**, i.e., the planes orthogonal to the principal directions.

5.2.3. Special Forms of the Stress Tensor.

We consider the stress tensor with respect to an orthonormal eigenbasis or another special basis, so that the corresponding matrix has a simpler form. We use the following terminology:

- **Uniaxial stress** for a stress tensor given by

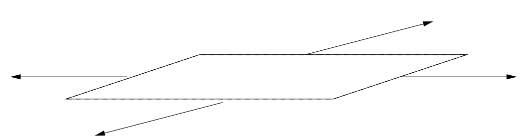
$$\begin{pmatrix} \sigma & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Example 5.15. This is the stress tensor in a long vertical rod loaded by hanging a weight on the end. □

- **Plane stressed state** or **biaxial stress** for a stress tensor given by

$$\begin{pmatrix} \sigma^{1} & 0 & 0 \\ 0 & \sigma^{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Example 5.16. This is the stress tensor in plate on which forces are applied as in the picture.



□

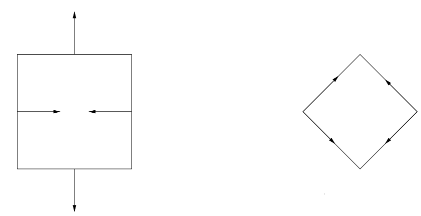
- **Pure shear** for a stress tensor given by

$$\begin{pmatrix} -\sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & \sigma & 0 \\ \sigma & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{5.8}$$

This is special case of the biaxial stress, in the case where $\sigma^1 = -\sigma^2$. In (5.8) the first is the stress tensor written with respect to an eigenbasis, while the second is the stress tensor written with respect to an orthonormal basis obtained by rotating an eigenbasis by $45°$ about the third axis. In fact

$$\begin{pmatrix} 0 & \sigma & 0 \\ \sigma & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{{}^{\mathrm{t}}L} \begin{pmatrix} -\sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{pmatrix} \underbrace{\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{L}$$

where L is the matrix of the change of coordinates.



- **Shear deformation** for a stress tensor given by

$$\begin{pmatrix} 0 & \sigma^{12} & \sigma^{13} \\ \sigma^{12} & 0 & \sigma^{23} \\ \sigma^{13} & \sigma^{23} & 0 \end{pmatrix}$$

(with respect to some orthonormal basis).

It turns out that the stress tensor σ is equivalent to a shear deformation if and only if its trace is zero.

EXAMPLE 5.17. The stress tensor

$$\begin{pmatrix} 2 & -4 & 0 \\ -4 & 0 & 4 \\ 0 & 4 & -2 \end{pmatrix}$$

represents a shear deformation. In fact, one can check that

$$\underbrace{\begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix}}_{{}^{\mathrm{t}}L} \begin{pmatrix} 2 & -4 & 0 \\ -4 & 0 & 4 \\ 0 & 4 & -2 \end{pmatrix} \underbrace{\begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix}}_{L} = \begin{pmatrix} 0 & 0 & -2 \\ 0 & 0 & 4\sqrt{2} \\ -2 & 4\sqrt{2} & 0 \end{pmatrix}$$

□

- **Hydrostatic pressure** with stress tensor given by

$$\begin{pmatrix} -p & 0 & 0 \\ 0 & -p & 0 \\ 0 & 0 & -p \end{pmatrix},$$

where $p \neq 0$ is the pressure. Here all eigenvalues are equal to $-p$.

EXAMPLE 5.18. Pressure of a fluid on a bubble. □

EXERCISE 5.19. Any stress tensor can be written as the sum of a hydrostatic pressure and a shear deformation. *Hint:* look at the next section.

5.2.4. Invariants.

Let A be a 3×3 matrix with entries a^{ij}. The *characteristic polynomial* $p_A(\lambda)$ of A is invariant under conjugation (see §1.4.2), so its coefficients remain unchanged.

$$\begin{aligned} p_A(\lambda) = &\det(A - \lambda I) = \det \begin{pmatrix} a^{11}-\lambda & a^{12} & a^{13} \\ a^{21} & a^{22}-\lambda & a^{23} \\ a^{31} & a^{32} & a^{33}-\lambda \end{pmatrix} \\ = &-\lambda^3 + \operatorname{tr} A \lambda^2 \\ &- \underbrace{\left(a^{11}a^{22} + a^{11}a^{33} + a^{22}a^{33} - a^{12}a^{21} - a^{23}a^{32} - a^{13}a^{31}\right)}_{\text{quadratic expression in the entries of } A} \lambda + \det A\,. \end{aligned}$$

Applying this to the stress tensor $\sigma = A$, we obtain some **stress invariants**, namely:

$$\begin{aligned} I_1 :=& \operatorname{tr}\sigma = \sigma^{11} + \sigma^{22} + \sigma^{33} \\ I_2 :=& (\sigma^{12})^2 + (\sigma^{23})^2 + (\sigma^{13})^2 - \sigma^{11}\sigma^{22} - \sigma^{22}\sigma^{33} - \sigma^{33}\sigma^{11} \\ I_3 :=& \det\sigma\,. \end{aligned}$$

That means, that the above quantities I_1, I_2 and I_3 are invariant when we change the orthonormal basis.[23]

The stress tensor can be expressed as the sum of 2 other stress tensors;

- The **hydrostatic stress tensor**

$$\pi\delta^{ij} = \begin{pmatrix} \pi & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & \pi \end{pmatrix},$$

where $\pi := I_1/3 = (\sigma^{11} + \sigma^{22} + \sigma^{33})/3$. This relates to *volume change*.

[23] By contravariance, when we change basis via a matrix L, the matrix of the stress tensor changes from σ to $\widetilde{\sigma} = {}^{t}\Lambda\sigma\Lambda$, where $\Lambda = L^{-1}$. But since we are restricting to orthonormal bases, we have that the change of basis matrix is orthogonal, i.e., $\Lambda = {}^{t}L$, so this is in fact a conjugation: $\widetilde{\sigma} = L\sigma L^{-1}$.

- The **deviatoric stress tensor**

$$s^{ij} := \sigma^{ij} - \pi\delta^{ij} = \begin{pmatrix} \sigma^{11} - \pi & \sigma^{12} & \sigma^{13} \\ \sigma^{12} & \sigma^{22} - \pi & \sigma^{23} \\ \sigma^{13} & \sigma^{23} & \sigma^{33} - \pi \end{pmatrix}.$$

  This relates to *shape change*.

REMARK 5.20. The **hydrostatic pressure** is generally defined as the negative one third of the stress invariant I_1, i.e. $p = -\pi$

REMARK 5.21. The stress deviator tensor is a shear deformation since

$$\mathrm{tr}(s^{ij}) = \sigma^{11} + \sigma^{22} + \sigma^{33} - 3\pi = 0.$$

Clearly

$$\sigma^{ij} = s^{ij} + \pi\delta^{ij}$$

and hence the stress tensor is a sum of a shear deformation and a hydrostatic pressure as claimed in Exercise 5.19.

5.2.5. Contravariance of the Stress Tensor.

Let $\mathcal{B} = \{e_1, e_2, e_3\}$ and $\widetilde{\mathcal{B}} = \{\tilde{e}_1, \tilde{e}_2, \tilde{e}_3\}$ be two orthonormal bases, and let

$$\tilde{e}_i = L_i^j e_j \qquad \text{and} \qquad e_i = \Lambda_i^j \tilde{e}_j\,, \tag{5.9}$$

where $L := L_{\widetilde{\mathcal{B}}\mathcal{B}}$ is the matrix of the change of basis and $\Lambda := L^{-1}$ is the inverse. Let $\mathbf{n}$ be a given unit vector and σ the stress across a surface perpendicular to $\mathbf{n}$. Then σ can be expressed in two ways, respectively with respect to $\mathcal{B}$ and to $\widetilde{\mathcal{B}}$,

$$\sigma = \sigma^{ij}(\mathbf{n}\cdot e_i)e_j \quad \text{and} \quad \sigma = \tilde{\sigma}^{ij}(\mathbf{n}\cdot \tilde{e}_i)\tilde{e}_j\,, \tag{5.10}$$

and we want to relate σ^{ij} to $\tilde{\sigma}^{ij}$. We start with the first expression for S in (5.10) and rename the indices for later convenience:

$$\sigma = \sigma^{km}(\mathbf{n}\cdot e_k)e_m = \sigma^{km}(\mathbf{n}\cdot \Lambda_k^i\tilde{e}_i)(\Lambda_m^j\tilde{e}_j) = \sigma^{km}\Lambda_k^i\Lambda_m^j(\mathbf{n}\cdot\tilde{e}_i)\tilde{e}_j\,,$$

where in the second equality we used (5.9), and in the third we used linearity. Comparing the last expression with the second expression in (5.10) we obtain

$$\tilde{\sigma}^{ij} = \sigma^{km}\Lambda_k^i\Lambda_m^j\,,$$

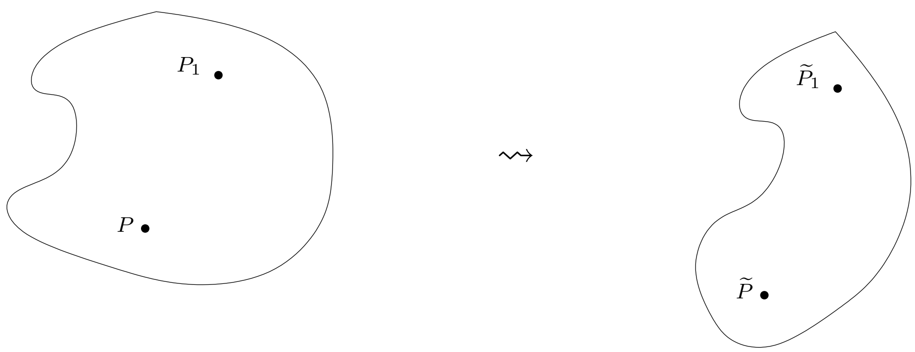
thus showing that σ is a **contravariant** 2-**tensor** or a **tensor of type** $(0,2)$.

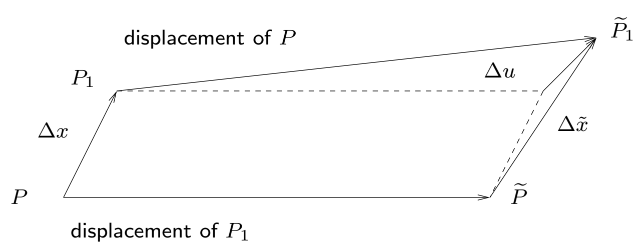
Heuristically, we may think that the stress at each point takes one plane, thought of as the kernel of a linear form, and gives a vector, thus is a linear map $V^* \to V$ or, equivalently, a bilinear map $V^* \times V^* \to \mathbb{R}$, i.e., a tensor of type $(0,2)$.

5.3. Strain Tensor (Verzerrung)

5.3.1. Physical Preliminaries.

Consider a slight deformation of a body, where we compare the relative positions of two particles before and after the deformation:





We have

$$\Delta \tilde{x} = \Delta x + \Delta u \,,$$

where Δx is the old relative position of P and P_1, $\Delta \tilde{x}$ is their new relative position and Δu is the displacement difference, which hence measures the deformation.

Assume that we have a small homogeneous deformation, that is

$$\Delta u = f(\Delta x) \,,$$

in other words, f is a small linear function independent of the point P. If we write the components of Δu and Δx with respect to an orthonormal basis $\{e_1, e_2, e_3\}$, the function f will be represented by a matrix with entries that we denote by f_{ij},

$$\Delta u_i = f_{ij} \Delta x^j \,.$$

The matrix (f_{ij}) can be written as a sum of a symmetric and an antisymmetric matrix as follows:

$$f_{ij} = \epsilon_{ij} + \omega_{ij}\,,$$

where

$$\epsilon_{ij} = \frac{1}{2}(f_{ij} + f_{ji})$$

is a symmetric matrix and is called the **strain tensor** or **deformation tensor** and

$$\omega_{ij} = \frac{1}{2}(f_{ij} - f_{ji})$$

is an antisymmetric matrix called the **rotation tensor**. We will next try to understand where these names come from.

REMARK 5.22. First we verify that a (small) antisymmetric 3×3 matrix represents a (small) rotation in 3-dimensional space.

FACT 5.23. *Let V be a vector space with orthonormal basis $\mathcal{B} = \{e_1, e_2, e_3\}$, and let $\omega = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$. The matrix R_ω of the linear map $V \to V$ defined by $v \mapsto \omega \times v$ with respect to the basis $\mathcal{B}$ is*

$$R_\omega = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}\,.$$

Indeed, we have

$$\begin{aligned} \omega \times v = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \det \begin{pmatrix} e_1 & e_2 & e_3 \\ a & b & c \\ x & y & z \end{pmatrix} \\ &= \begin{pmatrix} bz - cy \\ cx - az \\ ay - bx \end{pmatrix} = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}\,. \end{aligned}$$

Note that the matrix $R_\omega = (\omega_{ij}) := \begin{pmatrix} 0 & \omega_{12} & -\omega_{31} \\ -\omega_{12} & 0 & \omega_{23} \\ \omega_{31} & -\omega_{23} & 0 \end{pmatrix}$ corresponds to the cross product with the vector $\omega = \begin{pmatrix} -\omega_{23} \\ -\omega_{31} \\ -\omega_{12} \end{pmatrix}$. □

5.3.2. The Antisymmetric Case: Rotation.

Suppose that the matrix (f_{ij}) was already *antisymmetric*, so that

$$\omega_{ij} = f_{ij} \qquad \text{and} \qquad \epsilon_{ij} = 0\,.$$

Note that

$$\omega_{ii} = \frac{1}{2}(f_{ii} - f_{ii}) = 0.$$

By the Fact 5.23, if $\omega = \begin{pmatrix} -\omega_{23} \\ -\omega_{31} \\ -\omega_{12} \end{pmatrix}$, then

$$R_\omega \Delta x = \omega \times \Delta x$$

and the relation

$$\Delta u^i = f_{ij} \Delta x^j \tag{5.11}$$

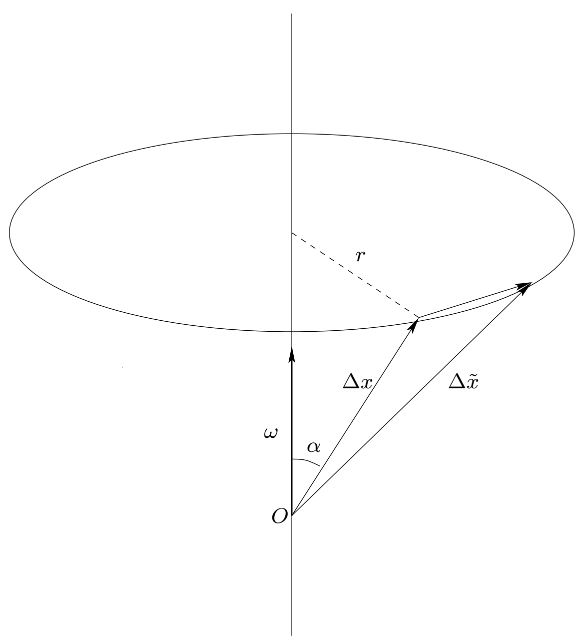
is equivalent to

$$\Delta u = \omega \times \Delta x\,,$$

so that

$$\Delta \tilde{x} = \Delta x + \Delta u = \Delta x + \omega \times \Delta x\,.$$

When ω is small, this represents an infinitesimal rotation of an angle $\|\omega\|$ about the axis $O\omega$.



In fact, since $\omega \times \Delta x$ is orthogonal to the plane determined by ω and by Δx, it is tangent to the circle with center along the axis $O\omega$ and radius determined by Δx. Moreover,

$$\|\Delta u\| = \|\omega \times \Delta x\| = \|\omega\| \underbrace{\|\Delta x\| \sin\alpha}_{r},$$

and hence, since the length of an arc of a circle of radius r corresponding to an angle θ is $r\theta$, infinitesimally this represents a rotation by an angle $\|\omega\|$.

5.3.3. The Symmetric Case: Strain.

The opposite extreme case is when the matrix f_{ij} was already *symmetric*, so that

$$\epsilon_{ij} = f_{ij} \qquad \text{and} \qquad \omega_{ij} = 0\,.$$

We will see that it is ϵ_{ij} that encodes the changes in the distances: in fact,

$$\begin{aligned} \|\Delta\tilde{x}\|^2 &= \Delta\tilde{x} \cdot \Delta\tilde{x} = (\Delta x + \Delta u) \cdot (\Delta x + \Delta u) \\ &= \Delta x \cdot \Delta x + 2\Delta x \cdot \Delta u + \Delta u \cdot \Delta u \\ &\simeq \|\Delta x\|^2 + 2\epsilon_{ij}\Delta x^i \Delta x^j\,, \end{aligned} \tag{5.12}$$

where in the last step we neglected the term $\|\Delta u\|^2$ since it is small compared to Δu when $\Delta u \to 0$ and used (5.11).

Remark 5.24. Even when f_{ij} is not purely symmetric, only the symmetric part ϵ_{ij} is relevant for the distortion of the distances. In fact, since ω_{ij} is antisymmetric, the term $2\omega_{ij}\Delta x^i \Delta x^j = 0$, so that

$$\|\Delta\tilde{x}\|^2 \simeq \|\Delta x\|^2 + 2f_{ij}\Delta x^i \Delta x^j = \|\Delta x\|^2 + 2\epsilon_{ij}\Delta x^i \Delta x^j\,.$$

□

Recall that a metric tensor (or inner product) encodes the distances among points. It follows that a deformation changes the metric tensor. Let us denote by g the metric before the deformation and by $\tilde{g}$ the metric after the deformation. By definition, we have

$$\|\Delta\tilde{x}\|^2 \overset{\text{def}}{=} \tilde{g}(\Delta\tilde{x}, \Delta\tilde{x}) = \tilde{g}_{ij}\Delta\tilde{x}^i \Delta\tilde{x}^j = \tilde{g}_{ij}(\Delta x^i + \Delta u^i)(\Delta x^j + \Delta u^j) \tag{5.13}$$

and

$$\|\Delta x\|^2 \overset{\text{def}}{=} g(\Delta x, \Delta x) = g_{ij}\Delta x^i \Delta x^j\,. \tag{5.14}$$

For infinitesimal deformations (that is, if $\Delta u \sim 0$), (5.13) becomes

$$\|\Delta\tilde{x}\|^2 = \tilde{g}_{ij}\Delta x^i \Delta x^j\,.$$

This, together with (5.14) and (5.12), leads to

$$\tilde{g}_{ij}\Delta x^i \Delta x^j \simeq g_{ij}\Delta x^i \Delta x^j + 2\epsilon_{ij}\Delta x^i \Delta x^j$$

and hence

$$\epsilon_{ij} \simeq \frac{1}{2}(\tilde{g}_{ij} - g_{ij}),$$

that is, ϵ_{ij} measures the change in the metric.

By definition the strain tensor ϵ_{ij} is symmetric

$$\mathcal{E} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{12} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{13} & \epsilon_{23} & \epsilon_{33} \end{pmatrix},$$

where the terms on the diagonal (in green) determine the elongation or the contraction of the body along the coordinate directions e_1, e_2, e_3, and the terms above the diagonal (in orange) are the **shear components** of the strain tensor; that is ϵ_{ij} is the movement of a line element parallel to Oe_j towards Oe_i. Since it is a symmetric tensor, it can be orthogonally diagonalized (cf. Theorem 3.9), so we can find an orthonormal basis with respect to which $\mathcal{E}$ is given by

$$\begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_3 \end{pmatrix},$$

The eigenvalues of $\mathcal{E}$ are the **principal coefficients** of the deformation and the eigenspaces are the **principal directions** of the deformation.

5.3.4. Special Forms of the Strain Tensor.

We use the following terminology:

(1) **Shear deformation** when $\mathcal{E}$ is traceless

$$\operatorname{tr} \mathcal{E} = \epsilon_{11} + \epsilon_{22} + \epsilon_{33} = 0.$$

(2) **Uniform compression** when the principal coefficients of $\mathcal{E}$ are equal (and nonzero)

$$\begin{pmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & k \end{pmatrix}$$

Exercise 5.25. Any strain tensor can be written as the sum of a uniform compression and a shear deformation.

5.4. Elasticity Tensor

The stress tensor represents an *external exertion* on the material, while the strain tensor represents the *material reaction* to that exertion. In crystallography these are called **field tensors** because they represent imposed conditions, opposed to **matter tensors**, that represents material properties.

Hooke's law says that, for small deformations, stress is related to strain by a matter tensor called **elasticity tensor** or **stiffness tensor** E:

$$\sigma^{ij} = E^{ijk\ell}\epsilon_{kl} \,,$$

while the tensor relating strain to stress is the **compliance tensor** S:

$$\epsilon_{k\ell} = S_{ijk\ell}\sigma^{ij} \,.$$

The elasticity tensor has order 4, and hence in 3-dimensional space it has $3^4 = 81$ components. Luckily, symmetry reduces the number of *independent* components for $E^{ijk\ell}$.

(1) **Minor symmetries:** The symmetry of the stress tensor

$$\sigma^{ij} = \sigma^{ji}$$

implies that

$$E^{ijk\ell} = E^{jik\ell} \qquad \text{for each } k, \ell \,;$$

it follows that for each k, ℓ fixed there are only 6 independent components $E^{ijk\ell}$

$$\begin{pmatrix} E^{11k\ell} & E^{12k\ell} & E^{13k\ell} \\ E^{12k\ell} & E^{22k\ell} & E^{23k\ell} \\ E^{13k\ell} & E^{23k\ell} & E^{33k\ell} \,. \end{pmatrix}$$

Having taken this in consideration, the number of independent components decreases to 6×3^2 at the most. Moreover, the symmetry also of the strain tensor

$$\epsilon_{k\ell} = \epsilon_{\ell k}$$

implies that

$$E^{ijk\ell} = E^{ij\ell k} \qquad \text{for each } i, j \,.$$

This means that for each i, j fixed there are also only 6 independent components $E^{ijk\ell}$, so that $E^{ijk\ell}$ has at most $6^2 = 36$ independent components.

(2) **Major symmetries:** Since (under appropriate conditions) partial derivatives commute, if follows from the existence of a *strain energy density functional* U satisfying

$$\frac{\partial^2 U}{\partial \epsilon_{ij} \partial \epsilon_{k\ell}} = E^{ijk\ell}$$

that

$$E^{ijk\ell} = E^{k\ell ij} \,,$$

that means the matrix with rows labelled by (i, j) and columns labelled by (k, ℓ) is symmetric. Since, from (1), there are only 6 entries (i, j) for a fixed (k, ℓ),

$E^{ijk\ell}$ can be written in a 6×6 matrix with rows labelled by (i, j) and columns labelled by (k, ℓ)

$$\begin{pmatrix} * & * & * & * & * & * \\ & * & * & * & * & * \\ & & * & * & * & * \\ & & & * & * & * \\ & & & & * & * \\ & & & & & * \end{pmatrix}$$

so that $E^{ijk\ell}$ has in fact only $6 + 5 + 4 + 3 + 2 + 1 = 21$ components.

5.5. Conductivity Tensor

Consider a homogeneous continuous crystal. Its properties can be divided into two classes:

- Properties that *do not depend* on a direction, and are hence described by *scalars*. Examples are density and heat capacity.
- Properties that *depend* on a direction, and are hence described by *tensors*. Examples are **elasticity**, **electrical conductivity** and **heat conductivity**. We say that a crystal is **anisotropic** when it has such *tensorial* properties.

5.5.1. Electrical Conductivity.

Let E be the **electric field** and J the **electrical current density**. We assume that these are constant throughout the crystal. At each point of the crystal:

(1) E gives the **electric force** (in Volts/m) that would be exerted on a positive test charge (of 1 Coulomb) placed at the point;
(2) J (in Amperes/m^2) gives the direction the charge carriers move and the **rate of electric current** across an infinitesimal surface perpendicular to that direction.

The electrical current density J is a function of the electric field E,

$$J = f(E)\,.$$

Consider a small increment ΔJ in J caused by a small increment ΔE in E, and write these increments in terms of their components with respect to a chosen orthonormal basis $\{e_1, e_2, e_3\}$.

$$\Delta J = \Delta J^i e_i \qquad \text{and} \qquad \Delta E = \Delta E^i e_i\,.$$

By Calculus, the increments are related by

$$\Delta J^i = \frac{\partial f^i}{\partial E^j}\Delta E^j + \text{higher order terms in } (\Delta E^j)^2, (\Delta E^j)^3, \ldots$$

If the quantities ΔE^j are small, we can assume that

$$\Delta J^i = \frac{\partial f^i}{\partial E^j}\Delta E^j \tag{5.15}$$

If we assume that $\frac{\partial f^i}{\partial E^j}$ is independent of the point of the crystal,

$$\frac{\partial f^i}{\partial E^j} = \kappa^i_j \in \mathbb{R}$$

we obtain the relation

$$\Delta J^i = \kappa^i_j \Delta E^j$$

or simply

$$\Delta J = \kappa\, \Delta E\,,$$

where κ if the **electrical conductivity tensor**. This is a $(1,1)$-tensor and may depend[24] on the initial value of E, that is the electrical conductivity may be different for small and large electric forces. If initially $E = 0$ and κ^0 is the corresponding electrical conductivity tensor, we obtain the relation

$$J = \kappa^0\, E$$

that is called the *generalized Ohm law*. This is always under the assumption that ΔE and ΔJ are small and that the relation is linear.

The **electrical resistivity tensor** is the inverse of κ:

$$\rho := \kappa^{-1}\,,$$

that is, it is the $(1,1)$-tensor such that

$$\rho^j_i \kappa^\ell_j = \delta^\ell_i\,.$$

The electrical conductivity measures the material's ability to conduct an electrical current, while the electrical resistivity quantifies the ability of the material to oppose the flow of the electrical current.

For an *isotropic* crystal, all directions are equivalent and these tensors are *spherical*, meaning

$$\kappa^j_i = k\delta^j_i \qquad \text{and} \qquad \rho^j_i = \frac{1}{k}\delta^j_i\,, \tag{5.16}$$

where k is a scalar, called the **electrical conductivity** of the crystal. Equation (5.16) can also be written as

$$\begin{pmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & k \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} \frac{1}{k} & 0 & 0 \\ 0 & \frac{1}{k} & 0 \\ 0 & 0 & \frac{1}{k} \end{pmatrix}\,.$$

[24]Typically, if the dependence between E and J is linear for any value, and not only for small ones, the tensor κ will not depend on the initial value of E.

In general, κ_i^j is neither symmetric nor antisymmetric (and actually *symmetry* does not even make sense for a $(1,1)$ tensor unless a metric is fixed, since it does require a canonical identification of V with V^*).

5.5.2. Heat Conductivity.

Let T be the *temperature* and H the *heat flux vector*. For a homogeneous crystal and constant H and for a constant gradient of T, *Fourier heat conduction law* says that

$$H = -K \operatorname{grad} T\,. \tag{5.17}$$

At each point of the crystal:

(1) $\operatorname{grad} T$ points in the direction of the highest ascent of the temperature and measures the rate of increase of T in that direction. The minus sign in (5.17) comes from the fact that the heat flows in the direction of decreasing temperature.
(2) H measure the amount of heat passing per unit area perpendicular to its direction per unit time.

Here, K is the **heat conductivity tensor** or **thermal conductivity tensor**. In terms of components with respect to a chosen orthonormal basis, we have

$$H^i = -K^{ij}(\operatorname{grad} T)_j\,.$$

Exercise 5.26. Verify that the gradient of a real function is a **covariant** 1-**tensor**.

The heat conductivity tensor is a contravariant 2-tensor and experiments show that it is symmetric and hence can be orthogonally diagonalized. The **heat resistivity tensor** is its inverse:

$$r := K^{-1}\,,$$

and hence is also symmetric. With respect to an orthonormal basis, K is represented by

$$\begin{pmatrix} K_1 & 0 & 0 \\ 0 & K_2 & 0 \\ 0 & 0 & K_3 \end{pmatrix}\,,$$

where the eigenvalues of K are called the **principal coefficients** of heat conductivity.

The fact that heat flows always in the direction of decreasing temperature shows that the eigenvalues are positive

$$K_i > 0\,.$$

The eigenspaces of K are called the **principal directions** of heat conductivity.

Solutions to Exercises

EXERCISE 1.4: (1) yes; (2) no, (3) no, (4) yes, (5) no.

EXERCISE 1.23:

(1) The vectors in $\mathcal{B}$ span V, since any element of V is of the form

$$\begin{bmatrix} a & b \\ c & -a \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} .$$

Moreover, the vectors in $\mathcal{B}$ are linearly independent since

$$a \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

if and only if

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

that is, if and only if $a = b = c = 0$.

(2) We know that $\dim V = 3$, as $\mathcal{B}$ is a basis of V and has three elements. Since $\widetilde{\mathcal{B}}$ also has three elements, it is enough to check either that it spans V or that it consists of linearly independent vectors. We will check this last condition. Indeed,

$$a \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + b \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \Longleftrightarrow \begin{bmatrix} a & c-b \\ b+c & -a \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

that is,

$$\begin{cases} a = 0 \\ b + c = 0 \\ c - b = 0 \end{cases} \Longleftrightarrow \begin{cases} a = 0 \\ b = 0 \\ c = 0. \end{cases}$$

(3) Since

$$\begin{bmatrix} 2 & 1 \\ 7 & -2 \end{bmatrix} = 2 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + 1 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 7 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} ,$$

we have

$$[v]_{\mathcal{B}} = \begin{pmatrix} 2 \\ 1 \\ 7 \end{pmatrix} .$$

To compute the coordinates of v with respect to $\widetilde{\mathcal{B}}$ we need to find $a, b, c \in \mathbb{R}$ such that

$$\begin{bmatrix} 2 & 1 \\ 7 & -2 \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + b \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} .$$

Solving the corresponding system of linear equations as above yields

$$[v]_{\widetilde{\mathcal{B}}} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} .$$

EXERCISE 2.2: (1) no; (2) no; (3) yes.

EXERCISE 2.3: (1) yes; (2) yes; (3) no.

EXERCISE 2.7: We show that V^* is a subspace of the vector space of all real-valued functions on V (cf. Example 1.3(3) on page 8), by checking the three conditions:

(1) The 0-function $\mathbf{0}$ associating the number zero to each vector in V is linear because $0 + 0 = 0$ and $k0 = 0$ for every $k \in \mathbb{R}$, so $\mathbf{0} \in V^*$;
(2) V^* is *closed under addition* since, if $\alpha : V \to \mathbb{R}$ and $\beta : V \to \mathbb{R}$ are linear, then $\alpha + \beta : V \to \mathbb{R}$ defined by $(\alpha + \beta)(v) = \alpha(v) + \beta(v)$ is also linear (in $v \in V$);
(3) V^* is *closed under multiplication by scalars* since, if $\alpha : V \to \mathbb{R}$ is linear and $k \in \mathbb{R}$, then $k\alpha : V \to \mathbb{R}$ defined by $(k\alpha)(v) = k(\alpha(v))$ is also linear.

EXERCISE 2.12: In fact:

$$\begin{aligned}
[\alpha]_{\mathcal{B}^*} L = \begin{pmatrix} \alpha_1 & \dots & \alpha_n \end{pmatrix} & \begin{bmatrix} L_1^1 & \dots & L_n^1 \\ \vdots & & \vdots \\ L_1^n & \dots & L_n^n \end{bmatrix} \\
&= \begin{pmatrix} \alpha_i L_1^i & \dots & \alpha_i L_n^i \end{pmatrix} \\
&= \begin{pmatrix} \widetilde{\alpha}_1 & \dots & \widetilde{\alpha}_n \end{pmatrix} \\
&= [\alpha]_{\widetilde{\mathcal{B}}^*}
\end{aligned}$$

EXERCISE 2.17: By the Leibniz formula, we have

$$\begin{aligned}
\det \begin{bmatrix} u \\ v \\ w \end{bmatrix} &= \det \begin{bmatrix} u^1 & u^2 & u^3 \\ v^1 & v^2 & v^3 \\ w^1 & w^2 & w^3 \end{bmatrix} \\
&= u^1 \left(v^2 w^3 - v^3 w^2 \right) + u^2 \left(v^3 w^1 - v^1 w^3 \right) + u^3 \left(v^1 w^2 - v^2 w^1 \right) \\
&= \begin{bmatrix} u^1 \\ u^2 \\ u^3 \end{bmatrix} \cdot \begin{bmatrix} v^2 w^3 - v^3 w^2 \\ v^3 w^1 - v^1 w^3 \\ v^1 w^2 - v^2 w^1 \end{bmatrix} = u \cdot (v \times w) .
\end{aligned}$$

EXERCISE 2.19: (1) yes; (2) yes; (3) yes; (4) no, because $v \times w$ is not a real number; (5) no, because it fails linearity (the area of the parallelogram spanned by v and w is the *same* as that of the parallelogram spanned by $-v$ and w); (6) no, because because it fails linearity in the second argument (the determinant of a matrix with $n > 1$ is not linear in that matrix).

EXERCISE 2.29: In fact,

$$\begin{aligned}
{}^{\mathrm{t}}LBL &= \begin{bmatrix} L_1^1 & \dots & L_1^n \\ \vdots & & \vdots \\ L_n^1 & \dots & L_n^n \end{bmatrix} \begin{bmatrix} B_{11} & \dots & B_{1n} \\ \vdots & & \vdots \\ B_{n1} & \dots & B_{nn} \end{bmatrix} \begin{bmatrix} L_1^1 & \dots & L_n^1 \\ \vdots & & \vdots \\ L_1^n & \dots & L_n^n \end{bmatrix} \\
&= \begin{bmatrix} L_1^1 & \dots & L_1^n \\ \vdots & & \vdots \\ L_n^1 & \dots & L_n^n \end{bmatrix} \begin{bmatrix} B_{1i}L_1^i & \dots & B_{1i}L_n^i \\ \vdots & & \vdots \\ B_{ni}L_1^i & \dots & B_{ni}L_n^i \end{bmatrix} \\
&= \begin{bmatrix} L_1^k B_{ki} L_1^i & \dots & L_1^k B_{ki} L_n^i \\ \vdots & & \vdots \\ L_n^k B_{ki} L_1^i & \dots & L_n^k B_{ki} L_n^i \end{bmatrix} \\
&= \widetilde{B} \,.
\end{aligned}$$

EXERCISE 3.3:

(1) no, as φ is *negative definite*, that is $\varphi(v, v) < 0$ if $v \in V$, $v \neq 0$;
(2) no, as φ is not symmetric;
(3) no, as φ is not positive definite;
(4) no, as φ is not positive definite;
(5) yes;
(6) yes.

EXERCISE 3.4:

(1) Yes, in fact:
   (a) $\int_0^1 p(x)q(x)dx = \int_0^1 q(x)p(x)dx$ because $p(x)q(x) = q(x)p(x)$;
   (b) $\int_0^1 (p(x))^2 dx \geq 0$ for all $p \in \mathbb{R}[x]_2$ because $(p(x))^2 \geq 0$, and $\int_0^1 (p(x))^2 dx = 0$ only when $p(x) = 0$ for all $x \in [0, 1]$, that is only if $p \equiv 0$.
(2) No, since $\int_0^1 (p'(x))^2 dx = 0$ implies that $p'(x) = 0$ for all $x \in [0, 1]$, but such p is not necessarily the zero polynomial.
(3) Yes.
(4) No. Is there $p \in \mathbb{R}[x]_2$, $p \neq 0$ such that $(p(1))^2 + (p(2))^2 = 0$?
(5) Yes. Is there a non-zero polynomial of degree 2 with 3 distinct zeros?
(6) No, since this is not symmetric.

EXERCISE 3.12: We write

$$[v]_{\widetilde{\mathcal{B}}} = \begin{pmatrix} \tilde{v}^1 \\ \tilde{v}^2 \\ \tilde{v}^3 \end{pmatrix} \quad \text{and} \quad [w]_{\widetilde{\mathcal{B}}} = \begin{pmatrix} \tilde{w}^1 \\ \tilde{w}^2 \\ \tilde{w}^3 \end{pmatrix}.$$

We know that g with respect to the basis $\widetilde{\mathcal{B}}$ has the standard form $g(v,w) = \tilde{v}^i \tilde{w}^i$ and we want to verify (3.7) using the matrix of the change of coordinates $L^{-1} = \Lambda$. If

$$[v]_{\mathcal{B}} = \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} \quad \text{and} \quad [w]_{\mathcal{B}} = \begin{pmatrix} w^1 \\ w^2 \\ w^3 \end{pmatrix}$$

then we have that

$$\begin{pmatrix} \tilde{v}^1 \\ \tilde{v}^2 \\ \tilde{v}^3 \end{pmatrix} = \Lambda \begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix} = \begin{pmatrix} v^1 - v^2 \\ v^2 - v^3 \\ v^3 \end{pmatrix}$$

and

$$\begin{pmatrix} \tilde{w}^1 \\ \tilde{w}^2 \\ \tilde{w}^3 \end{pmatrix} = \Lambda \begin{pmatrix} w^1 \\ w^2 \\ w^3 \end{pmatrix} = \begin{pmatrix} w^1 - w^2 \\ w^2 - w^3 \\ w^3 \end{pmatrix}$$

It follows that

$$\begin{aligned} g(v,w) = \tilde{v}^i \tilde{w}^i &= (v^1 - v^2)(w^1 - w^2) + (v^2 - v^3)(w^2 - w^3) + v^3 w^3 \\ &= v^1 w^1 - v^1 w^2 - v^2 w^1 + 2v^2 w^2 - v^2 w^3 - w^3 v^2 + 2v^3 w^3 . \end{aligned}$$

EXERCISE 3.15: With respect to $\widetilde{\mathcal{B}}$, we have

$$\begin{aligned} \|v\| &= (1^2 + 1^2 + 1^2)^{1/2} = \sqrt{3} \\ \|w\| &= ((-1)^2 + (-1)^2 + 3^2)^{1/2} = \sqrt{11} \end{aligned}$$

and with respect to $\mathcal{E}$

$$\begin{aligned} \|v\| &= (3 \cdot 3 - 3 \cdot 2 - 2 \cdot 3 + 2 \cdot 2 \cdot 2 - 2 \cdot 1 - 1 \cdot 2 + 2 \cdot 1 \cdot 1)^{1/2} = \sqrt{3} \\ \|w\| &= (1 \cdot 1 - 1 \cdot 2 - 2 \cdot 1 + 2 \cdot 2 \cdot 2 - 2 \cdot 3 - 3 \cdot 2 + 2 \cdot 3 \cdot 3)^{1/2} = \sqrt{11} . \end{aligned}$$

EXERCISE 3.16: Saying that the orthogonality is meant with respect to g, means that we have to show that $g(v - \operatorname{proj}_{b_k} v, b_k) = 0$. In fact,

$$g(v - \operatorname{proj}_{b_k} v, b_k) = g(v - \frac{g(v,b_k)}{g(b_k,b_k)} b_k, b_k) = g(v,b_k) - \frac{g(v,b_k)}{\cancel{g(b_k,b_k)}} \cancel{g(b_k,b_k)} = 0$$

EXERCISE 3.20:

(1) The coordinate vectors of basis vectors with respect to that same basis are simply standard vectors, in this case:
$$[\tilde{b}_1]_{\widetilde{\mathcal{B}}} = \begin{pmatrix}1\\0\\0\end{pmatrix},\ [\tilde{b}_2]_{\widetilde{\mathcal{B}}} = \begin{pmatrix}0\\1\\0\end{pmatrix} \text{ and } [\tilde{b}_3]_{\widetilde{\mathcal{B}}} = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

(2) As in Example 3.11, we have $G_{\widetilde{\mathcal{B}}} = I$ and $G_{\mathcal{E}} = \begin{bmatrix}1 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 2\end{bmatrix}$.

(3) In parts (a) and (b), note that, for an orthonormal basis $\widetilde{\mathcal{B}}$, we have $\widetilde{\mathcal{B}}^g = \widetilde{\mathcal{B}}$. In parts (c) and (d), we use the computations in Example 3.19 and the fact that $[v]_{\mathcal{E}} = L_{\widetilde{\mathcal{B}}\mathcal{E}}[v]_{\widetilde{\mathcal{B}}}$.

(a) $[\tilde{b}^1]_{\widetilde{\mathcal{B}}} = [\tilde{b}_1]_{\widetilde{\mathcal{B}}} = \begin{pmatrix}1\\0\\0\end{pmatrix}$, $[\tilde{b}^2]_{\widetilde{\mathcal{B}}} = [\tilde{b}_2]_{\widetilde{\mathcal{B}}} = \begin{pmatrix}0\\1\\0\end{pmatrix}$ and $[\tilde{b}^3]_{\widetilde{\mathcal{B}}} = [\tilde{b}_3]_{\widetilde{\mathcal{B}}} = \begin{pmatrix}0\\0\\1\end{pmatrix}$.

(b) $[\tilde{b}^1]_{\mathcal{E}} = [\tilde{b}_1]_{\mathcal{E}} = \begin{pmatrix}1\\0\\0\end{pmatrix}$, $[\tilde{b}^2]_{\mathcal{E}} = [\tilde{b}_2]_{\mathcal{E}} = \begin{pmatrix}1\\1\\0\end{pmatrix}$ and $[\tilde{b}^3]_{\mathcal{E}} = [\tilde{b}_3]_{\mathcal{E}} = \begin{pmatrix}1\\1\\1\end{pmatrix}$.

(c) $[e^1]_{\widetilde{\mathcal{B}}} = \begin{pmatrix}1\\1\\1\end{pmatrix}$, $[e^2]_{\widetilde{\mathcal{B}}} = \begin{pmatrix}0\\1\\1\end{pmatrix}$ and $[e^3]_{\widetilde{\mathcal{B}}} = \begin{pmatrix}0\\0\\1\end{pmatrix}$.

(d) $[e^1]_{\mathcal{E}} = \begin{pmatrix}3\\2\\1\end{pmatrix}$, $[e^2]_{\mathcal{E}} = \begin{pmatrix}2\\2\\1\end{pmatrix}$ and $[e^3]_{\mathcal{E}} = \begin{pmatrix}1\\1\\1\end{pmatrix}$.

EXERCISE 3.23:

(1) The assertion in the case of the bases $\mathcal{E}$ and $\mathcal{E}^g$ follows from
$$G_{\mathcal{E}}^{-1} = (L_{\mathcal{E}\mathcal{E}^g})^{-1} = \begin{bmatrix}| & | & |\\ e^1 & e^2 & e^3\\ | & | & |\end{bmatrix} = \begin{bmatrix}3 & 2 & 1\\ 2 & 2 & 1\\ 1 & 1 & 1\end{bmatrix}.$$

(2) Since $\tilde{b}^j = \tilde{b}_j$, we have $G_{\widetilde{\mathcal{B}}}^{-1} = L_{\widetilde{\mathcal{B}}\widetilde{\mathcal{B}}^g} = I$ and (3.15) is immediately verified.

EXERCISE 4.7: We first use upper indices for rows and lower indices for columns and Assume that A and B are square matrices of the same size. If A has (i,j)-entry (where i labels the row and j the column) A^i_j and B has (i,j)-entry B^i_j, then by the definition of matrix product the matrix $C := AB$ has (i,j)-entry
$$A^i_k B^k_j$$
and the transpose of A has (i,j)-entry A^j_i, so $C^{\mathrm{t}}A$ has (i,j)-entry
$$C^i_\ell A^j_\ell$$

and $AB^{\mathsf{t}}A$ has (i,j)-entry

$$A^i_k B^k_\ell A^j_\ell \,.$$

We obtain the formula (4.3) by replacing A by Λ and B^k_ℓ by $S^{k\ell}$.

EXERCISE 4.11: The coordinates of $b_1 \otimes a_1 + b_2 \otimes a_2$ with respect to the basis $b_i \otimes a_j$ are δ^{ij}. So the task amounts to showing that there is no solution to the equation

$$\begin{pmatrix} v^1 \\ v^2 \end{pmatrix} \begin{pmatrix} w^1 & w^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

where the v^i and w^j are the coordinates of arbitrary vectors $v \in V$ and $w \in W$ w.r.t. the given bases. Indeed, the system

$$\begin{cases} v^1 w^1 = 1 \\ v^1 w^2 = 0 \\ v^2 w^1 = 0 \\ v^2 w^2 = 1 \end{cases}$$

has no solution, since the first two equations force that w^2 be zero, yet this is impossible for the last equation.

EXERCISE 5.4: We choose a coordinate system with origin at the vertex O, with x-axis along the side of length a, y-axis along the side of length b and z-axis perpendicular to the plate. We already know that the mass density is constant equal to $\rho = \frac{m}{ab}$. Then

$$\begin{aligned} \underbrace{I_{11}}_{I_{xx}} &= \int_0^a \int_0^b (y^2 + \underbrace{z^2}_{=0}) \underbrace{\rho}_{\frac{m}{ab}} \, dy \, dx \\ &= \frac{m}{ab} a \int_0^b y^2 \, dy \\ &= \frac{m}{b} \left[\frac{y^3}{3} \right]_0^b \;=\; \frac{m}{3} b^2 \,. \end{aligned}$$

Similarly,

$$\underbrace{I_{22}}_{I_{yy}} = \frac{m}{3} a^2 \,,$$

and

$$\underbrace{I_{33}}_{I_{zz}} = \int_0^a \int_0^b (x^2 + y^2) \rho \, dy \, dx = \frac{m}{3}(a^2 + b^2)$$

is again just the sum of I_{11} and I_{22}.

Furthermore,

$$I_{23} = I_{32} = -\int_0^a \int_0^b y \underbrace{z}_{=0} \rho \, dy \, dx = 0 \,,$$

and, similarly, $I_{31} = I_{13} = 0$. Finally, we have

$$I_{21} = I_{21} = -\int_0^a \int_0^b xy \, \rho \, dy \, dx = -\frac{m}{ab} \left[\frac{x^2}{2}\right]_0^a \left[\frac{y^2}{2}\right]_0^b = -\frac{mab}{4} \,.$$

We conclude that the inertia tensor is given by the matrix

$$\tfrac{m}{12} \begin{pmatrix} 4b^2 & -3ab & 0 \\ -3ab & 4a^2 & 0 \\ 0 & 0 & 4(a^2+b^2) \end{pmatrix} .$$

Exercise 5.6: We again choose the unit vector

$$p = \begin{pmatrix} \frac{a}{\sqrt{a^2+b^2}} \\ \frac{b}{\sqrt{a^2+b^2}} \\ 0 \end{pmatrix} = \tfrac{1}{\sqrt{a^2+b^2}} \begin{pmatrix} a \\ b \\ 0 \end{pmatrix}$$

defining the axis of rotation, but now use the inertia tensor computed in Exercise 5.4,

$$\tfrac{m}{12} \begin{pmatrix} 4b^2 & -3ab & 0 \\ -3ab & 4a^2 & 0 \\ 0 & 0 & 4(a^2+b^2) \end{pmatrix} ,$$

thus obtaining

$$\begin{aligned} I_p =& I_{ij} p^i p^j \\ =& \tfrac{m}{12(a^2+b^2)} \begin{pmatrix} a & b & 0 \end{pmatrix} \begin{pmatrix} 4b^2 & -3ab & 0 \\ -3ab & 4a^2 & 0 \\ 0 & 0 & 4(a^2+b^2) \end{pmatrix} \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} \\ =& \tfrac{m}{6} \tfrac{a^2 b^2}{a^2+b^2} \,, \end{aligned}$$

necessarily the same result as in Example 5.5.

Exercise 5.7: We use the inertia tensor calculated in Example 5.3, where the origin of the coordinate system is at the center of mass, and choose $p = e_3$. The moment of inertia is then

$$\begin{aligned} I_p =& I_{ij} p^i p^j \\ =& \tfrac{m}{12} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b^2 & 0 & 0 \\ 0 & a^2 & 0 \\ 0 & 0 & a^2+b^2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ =& I_{33} = \tfrac{m}{12}(a^2+b^2) \,. \end{aligned}$$

EXERCISE 5.8: We use the inertia tensor calculated in Exercise 5.4, where the origin of the coordinate system is at a vertex of the plate, and choose $p = e_3$. The moment of inertia is then

$$\begin{aligned} I_p =& I_{ij} p^i p^j \\ =& \tfrac{m}{12} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4b^2 & -3ab & 0 \\ -3ab & 4a^2 & 0 \\ 0 & 0 & 4(a^2+b^2) \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ =& I_{33} = \tfrac{m}{3}(a^2+b^2)\,. \end{aligned}$$