Chapter 1

Confusions: What Are Tensors Exactly?

One way to learn a lot of mathematics is by reading the first chapters of many books.

— *Paul R. Halmos*

Tensors have profound applications in physics, computer science, engineering, machine learning, data mining, medicine (diffusion tensor imaging), etc. This chapter provides a background overview of tensors. You may find usage of terms that have not yet been defined. The purpose is to have a "big picture".

If you find the first chapter helpful, you might consider reading beyond it. The logical exposition starts in Chap. 3.

§1. Questions and Confusions

The concept of tensor is confusing to many students. If one does a search on the Internet, he can find many questions asked about tensors. For example:

> Is a tensor just a (higher dimensional) matrix?
> How long have tensors been around, and why is there a sudden fascination for tensors in machine learning?
> Are tensors in machine learning the same thing as tensors in mathematics and physics?
> Are tensors in machine learning contravariant or covariant?
> What is a metric tensor?
> Why is inertia tensor a tensor? (It is defined as a matrix in most of the books.)
> What is an example of a quantity that has the correct number of components but fails to be a tensor?
> What is the connection between tensor and tensor product?
> What is the physical meaning of a tensor?
> Can you add the components of a contravariant tensor and a covariant tensor?
> Do pure mathematicians have an interest in tensor analysis?
> What are some open problems in tensor analysis?
> Is tensor analysis relevant to deep learning?

There are many answers and explanations floating on the Internet. However, instead of solving the mysteries, many of these only add more confusion to the already confused learners. The following are a few examples:

> "A tensor is just an n-dimensional array with n indices."
>
> "Tensors are simply mathematical objects that can be used to describe physical properties."
>
> "Tensors are generalizations of scalars and vectors."
>
> "Basically tensors are vectors which have not a single direction but they rather point in all directions."
>
> "If I ask you what a vector is, you may tell me that is an element of a vector space, so tensor is an element of a tensor space."

> "Tensors have properties of both vectors and scalars, like area, stress etc."
>
> "A tensor is not a scalar, a vector or anything. It's just an abstract quantity that obeys the coordinate transformation law. Anything that satisfies the law is a tensor. That's it!"
>
> "In mathematics, tensors are geometrical objects that describe the linear relationships between geometric, numerical, and other tensile vectors."
>
> "The simplest way to imagine a tensor is that it's a vector in a product space. Each index denotes a factor of the product space in which the tensor lives, and may be raised or lowered depending on how the corresponding factor transforms under a change of basis. The number of indices counts the rank of a tensor. As such, tensors are essentially just generalizations of vectors. Their components (in a certain basis) are multidimensional arrays. A tensor is more than simply a multidimensional array, for the same reason that a vector is not simply a list of its components."
>
> "Speaking somewhat non-technically, tensors represent a linear operator of other tensors. Each time you operate a tensor on another tensor a set of matching indices disappears."
>
> "A tensor is a multilinear function."
>
> "A tensor, with the possibility of a multitude of indices, both covariant and contravariant, look like multidimensional data in 0, 1, 2, 3, and higher dimensions."
>
> "In the simplest form: the quantity having magnitude, direction and plane to act are called tensor quantities."
>
> "A tensor is an element of a tensor product of two or more vector spaces."
>
> "A tensor is the tensor product of two vectors."
>
> "Tensor: it is those physical quantity which may have tension-like effects."

Well, each of them speaks some truth about tensors, but they also reflect a lot of confusions. This reminds me of reading some funny answers of young children to the question "What is love".

* **Comparison:** What do love and tensor have in common?

"What is love?"

> "Love is when a girl puts on perfume and a boy puts on shaving cologne and they go out and smell each other." (age 5)
>
> "Love is when you tell a guy you like his shirt, then he wears it every day." (age 7)
>
> "If you want to learn to love better, you should start with a friend who you hate." (age 6)
>
> "Love is when mommy sees daddy smelly and sweaty and still says he is handsomer than Robert Redford." (age 8)
>
> "Love is when your puppy licks your face even after you left him alone all day." (age 4)
>
> "Love is when you kiss all the time. Then when you get tired of kissing, you still want to be together and you talk more." (age 8)
>
> "I know my older sister loves me because she gives me all her old clothes and has to go out and buy new ones." (age 4)
>
> "I let my big sister pick on me because my mom says she only picks on me because she loves me. So I pick on my baby sister because I love her." (age 4)

Each of these answers certainly tells some aspect of the truth.

What do love and tensor have in common? Is the love between sisters the same as that between mom and dad, dating teenagers, and dogs and humans? Compare with the question: is the tensor in machine learning the same as those in mathematics and physics?

The concept of love is abstract and complex, and it has never been rigorously defined. The tensor is also abstract and complex. It was poorly defined in the past. There are rigorous modern definitions, but at a cost of being more abstract and less intuitive. So the old-fashioned definition is hard to understand because it is not rigorous; the modern definition is hard to understand because it is rigorous. It is the goal of this book to explain the rigorous definitions of tensor in an intuitive way, so that students no longer have to recite those definitions like a parrot.

We shall have answers to these questions through this book. After reading the book, the reader should be able to judge the above quoted answers, which is correct and which is wrong. However, readers would like to have some quick answers before committing to reading a book. That is the purpose of this chapter.

§2. Who Invented the Tensor?

In this section, we give a brief history of the concept of tensor. This answers the question how long tensors have been around. It also answers the question "why are tensors confusing" from one perspective: it has different origins and it is the merge of different threads in history. In the next section we provide answers to this question from another aspect: there are many apparently different definitions of tensor in the current literature.

There were several threads in the development of tensor theory in late 1800s and early 1900s, including Ricci, Gibbs, Voigt and Whitney. Most modern authors give credit to Ricci for the concept of tensor, because the early textbooks, especially the physics literature, predominantly followed his definitions. Ricci did not use "tensor" in his definition, but rather "system". Physicists transplanted the name "tensor" to Ricci's definition. Although being called a "tensor", Ricci's definition actually defines a tensor field. This causes the most confusion to the beginners. Gibbs, Voigt and Whitney defined a tensor as a tensor in the algebraic sense.

(1) G. Ricci [(1892)]: covariant and contravariant systems, but he called those "systems", rather than "tensors" (what he defined is a tensor field in the modern sense; see more in Sec. 3).

(2) J. W. Gibbs [(1884)]: dyadics and polyadics (these are actually tensors in the modern sense, only by different names; see more in Chap. 4).

(3) W. Voigt [(1898)]: coined the name tensor—in a narrower sense of symmetric tensors in the study of elasticity of crystals.

(4) H. Whitney [(1937)]: tensor product (see more in Chap. 5).

Gibbs is recognized as one of the founders of vector algebra and vector analysis. Gibbs played an important role in emancipating vectors from Hamilton's quaternions. What is often underappreciated is his major contribution in the development of tensor algebra and tensor analysis (in

Euclidean space). Gibbs developed the concept of dyadics and polyadics. These are actually tensors in the modern sense, only by different names.[1] His dyadic product is exactly the tensor product in the modern sense, except his notation is the juxtaposition of two vectors $\mathbf{uv}$, compared with the modern notation of $\mathbf{u} \otimes \mathbf{v}$.

W. Voigt [(1898)] introduced the term tensor, in his study of stress and strain of crystals in his book *The Fundamental Physical Properties of the Crystals* (*Die fundamentalen physikalischen Eigenschaften der Krystallen*). The word "tensor" has its root "tensus" in Latin, meaning stretch or tension. Both stress and strain tensors are symmetric tensors of the second order and each has six components. Voigt denotes them as a 6-dimensional vector. This is known as the Voigt notation. The term tensor was adopted by physicists Max Abraham (1904), Arnold Sommerfeld (1910), Max von Laue (1911). Einstein and Grossmann [(1913)] [2] used Ricci's definition but with the name "tensor" instead of Ricci's name "system".

Whitney [(1937)] defined the tensor product. It is actually the idea of Gibbs dyadics made more precise. There are also other threads that are related to the development of tensors. Grassmann developed exterior algebra in 1862. Although exterior algebra can be established independent of the tensor theory, there is a connection between these two. An exterior vector is in fact an antisymmetric tensor. H. Minkowski [(1908)] introduced the electromagnetic tensor, which is an antisymmetric tensor, although he called it a "vector of the second kind" (of 6 dimensions, to distinguish it from a "vector of the first kind" with 4 dimensions). A. Sommerfeld later called it a 6-vector. Let us compare it with Voigt's tensor for stress, which is also expressed as a 6-vector. Voigt's tensor is a symmetric tensor over a 3-dimensional vector space, while the electromagnetic field tensor is an antisymmetric tensor over a 4-dimensional vector space.

Chap. 9 discusses the electromagnetic field tensor.

[1] The term tensor did appear in Gibbs' book, but was used to refer to a special type of tensors (namely a special type of linear transformations). W. R. Hamilton also used the term tensor, but referring to the modulus of a quaternion, which is totally irrelevant to our tensor theory. Tensor in Hamilton's sense is no longer in use today. Rather, it is called the modulus or norm of the quaternion.

[2] This paper has two parts put together, with Einstein as the single author for the physics part and Grossmann as the single author for the mathematics part.

* **Philosophical View**: Is mathematics invented or discovered?
—My opinion: It is both.

We asked the question "who invented the tensor". Was the tensor invented, or discovered? There is even an age-long philosophical question: "Is mathematics invented, or discovered?"

We asked the question "what is a tensor". In fact, a tensor is whatever we define it to be. We do have the liberty when it comes to definitions. In this sense, mathematics is an invention. Sherman Stein [(2010)] wrote a book, *Mathematics: the Man-made Universe*. The title of the book reflects this view. Of course, other people have argued that mathematics is discovery and this topic has been an unresolved debate.

My opinion is: it is both. In mathematics, we first invent this man-made universe. Then we make discoveries inside it. This man-made universe can be extremely complex and discovery in it is by no means a trivial process. For instance, the creation of non-Euclidean geometry is an invention, but its interpretations (or models) are discoveries, which uncover the connection between non-Euclidean and Euclidean geometries. Take group as another example. The definition of a group takes only a few lines of text, which can be viewed as an invention. The culminating result in group theory, the classification of the finite simple groups is a discovery, with tens of thousands of pages in several hundred articles written by about 100 authors, published mostly between 1955 and 2004. Riemannian manifold can be another example. Its definition also consists of just a few lines of text. The Nash embedding theorem is a great discovery, which reveals that although Riemannian manifold is defined intrinsically, it is always isometric to some submanifold embedded in some higher dimensional Euclidean space.

I have interpreted discovery as the discovery in the man-made universe of mathematics itself. Is mathematics about discovery in nature? My answer is yes and no: *no* in the sense that modern mathematics in its abstract form is liberated from the obligation of discovering the truth in nature, but *yes* in the sense that mathematics may be part of the process of discovering nature when it is applied in science. In the old days, mathematics was intended to discover the truth in nature directly, but in modern days, its participation in the discovery is indirect. Whatever abstract mathematics can be applied to the real world, if we find a *physical model* of the abstract *mathematical structure* (Appendix 2).

§3. Different Definitions of the Tensor

Why is the concept of tensor confusing? It is just a definition, isn't it? Think about the definition of an equilateral triangle. No one would have difficulty with that.

Some factors may make a concept hard to understand:

(1) The concept itself is more complex.

(2) The definition itself is not clear. Oftentimes the lack of rigor in the definition is caused by the intrinsic complexity of the concept itself. Historically, the first attempts to define a concept were often not successful in pinning down the essence of the concept. It may take centuries for the concept to evolve and get crystallized. Mathematics is full of evolution history of such concepts: complex numbers, real numbers, limit, continuity, vectors, ..., and the list goes on and on (see the boxes at the end of the section).

(3) Different definitions coexist in the literature, also due to historical reasons. Some of these definitions are equivalent, but not all of them are equivalent.

It turns out that all these factors have an effect on the concept of tensor. They cause many confusions for the beginners. In the following, we list several definitions of tensors that can be found in textbooks. Don't worry if you are confused with these. It is just to show that you do have a good reason to be confused, which is not your fault.

Definitions 1 and 2 are mostly seen in older textbooks of tensor analysis, physics, and especially general relativity.

Definition 1. A set of quantities ξ^{rs} is said to be a **contravariant tensor** (of degree 2) if under the change of coordinates

$$x'^i = x'^i(x^1, \ldots, x^n), \qquad i = 1, \ldots, n, \tag{1.1}$$

they transform according to

$$(\xi')^{st} = \sum_{\sigma,\tau} \xi^{\sigma\tau} \frac{\partial x'^s}{\partial x^\sigma} \frac{\partial x'^t}{\partial x^\tau}. \tag{1.2}$$

A set of quantities ξ_{lm} is said to be a **covariant tensor** if they transform according to

$$(\xi')_{lm} = \sum_{\lambda,\mu} \xi_{\lambda\mu} \frac{\partial x^\lambda}{\partial x'^l} \frac{\partial x^\mu}{\partial x'^m}. \tag{1.3}$$

A set of quantities ${\xi_l}^s$ is said to be a **mixed tensor** if they transform according to

$$(\xi')_l{}^s = \sum_{\lambda,\sigma} {\xi_\lambda}^\sigma \frac{\partial x^\lambda}{\partial x'^l} \frac{\partial x'^s}{\partial x^\sigma}. \tag{1.4}$$

Remark. This definition is basically due to Ricci. It is confusing that most books call these tensors, but what Ricci defines here are actually tensor fields. Ricci should not be blamed because he called these "systems". It is the use of the name tensor [Einstein and Grossmann (1913)] that causes the confusion of tensors with tensor fields. Each "quantity", or component ξ^{rs} is actually a function of space locations $\mathbf{x} = (x_1, \ldots, x_n)$. If the set of quantities is considered a single tensor ξ, then Ricci defines a tensor field $\xi(\mathbf{x})$, which is the assignment of a tensor ξ to each space point $\mathbf{x}$. A tensor ξ should be a single algebraic entity. Logically, a tensor as an algebraic entity should be defined first, before the definition of a tensor field, but this was not done by Ricci. This is the reason why Ricci used the components in his definition but amended by the coordinate transformation laws. In the modern perspective, these transformation laws are not necessary. They are the consequence of the basis change in the tangent space of the differentiable manifold, induced by local coordinate change Eq. 1.1 (see Sec. 3 in Chap. 10).

The arbitrary coordinate transformation Eq. 1.1 and the involvement of partial derivatives in the above definition clearly hint the tensor field. To make a seemingly algebraic definition of tensor, the general coordinate transformation Eq. 1.1 is restricted to linear transformations. This results in the following shy version of the definition.

Definition 2. A set of quantities ξ^{rs} is said to be a **contravariant tensor** (of degree 2) if under the change of coordinates

$$x'^i = \sum_k {\Lambda_k}^i x^k \tag{1.5}$$

and its inverse

$$x^k = \sum_i {\bar{\Lambda}_i}{}^k x'^i, \tag{1.6}$$

where the constant coefficients ${\Lambda_k}^i$ and ${\bar{\Lambda}_i}{}^k$ satisfy

$$\sum_r {\Lambda_i}^r {\bar{\Lambda}_r}{}^k = {\delta_i}^k, \tag{1.7}$$

they transform according to

$$(\xi')^{st} = \sum_{\sigma,\tau} \xi^{\sigma\tau} \Lambda_{\sigma}{}^{s} \Lambda_{\tau}{}^{t}. \tag{1.8}$$

A set of quantities ξ_{lm} is said to be a **covariant tensor** if they transform according to

$$(\xi')_{lm} = \sum_{\lambda,\mu} \xi_{\lambda\mu} \bar{\Lambda}_{l}{}^{\lambda} \bar{\Lambda}_{m}^{\mu}. \tag{1.9}$$

A set of quantities $\xi_l{}^s$ is said to be a **mixed tensor** if they transform according to

$$(\xi')_l{}^s = \sum_{\lambda,\sigma} \xi_{\lambda}{}^{\sigma} \bar{\Lambda}_{l}{}^{\lambda} \Lambda_{\sigma}{}^{s}. \tag{1.10}$$

Remark. Although this version looks more algebraic, the meaning of the linear coordinate transformation Eq. 1.5 is still not clear, if the set of quantities is an individual tensor instead of a tensor field. Furthermore, the meanings of "contravariant" and "covariant" are not apparent. According to K. Reich [(1994)], J. Sylvester introduced the terms "covariant" and "contravariant" in 1851 [Sylvester (1851)]. We shall reveal this in Sec. 2 of Chap. 6, these coordinate changes are with respect to the basis change of the underlying vector space, which involves a matrix $A_i{}^k$. Eq. 1.7 tells us that $\bar{\Lambda}_i{}^k$ is the transpose of the inverse of Λ_k^i. The matrix $\bar{\Lambda}_i{}^k$ here is same as $A_i{}^k$ in Sec. 2 of Chap. 6. That is why the transformation of covariant tensor involves $\bar{\Lambda}_i{}^k$, which means "the same as", or "together with" the transformation of the basis, while the contravariant tensor involves Λ_k^i, which is the inverse of $A_i{}^k$ with a meaning "against". We may call the basis transformation the "forward" transformation and its inverse the "backward" transformation. If the basis undergoes a forward transformation, the coordinates will undergo a "backward" transformation, as in Eq. 1.5, with an analogy: if the train moves forward, the trees outside seem to move in the backward direction from the perspective of someone inside the train. So the transformation for contravariant tensors is really "contra" to the basis transformation, which is not explicit here. It is rather "together with" the coordinate transformation of vectors Eq. 1.5. Eq. 1.5 itself is considered "contra", or "backward", with respect to the basis transformation. Another word of caution for the beginners is the popular tensor component notation in literature. Although $\bar{\Lambda}$ looks similar to Λ, it is actually the transpose of the inverse matrix of Λ.

g^{ij} are the components of the inverse matrix of the metric matrix g_{ij}.

This kind of definition of tensor is often referred to as the old-fashioned definition. It is this component approach that caused the conundrum, with the concept of tensor portrayed as an equivocal duality of matrix and non-matrix, just like the mixture of the living and the dead states of Schrödinger's cat. The tensor is defined as a matrix, but amended by the transformation laws. It is defined as the components of an object, without a clear definition of what this object is.

In recent years, with the booming research in machine learning, the machine learning community uses the tensor simply in the sense of a multi-dimensional array (or higher dimensional matrix), ignoring the transformation laws and breaking up this fuzzy duality. We shall discuss tensors in machine learning in Chap. 2.

Definition 3. (in the context of machine learning) A tensor is a multi-dimensional array (or matrix).

It is a trend in recent physics textbooks to use the following definition of a tensor.

Definition 4. Let V be a vector space over $\mathbb{R}$ and V^* be its dual space. A multilinear mapping

$$\Phi : \underbrace{V^* \times \cdots \times V^*}_{\text{p}} \times \underbrace{V \times \cdots \times V}_{\text{q}} \to \mathbb{R}$$

is called a tensor of type (p, q).

Remark. A question from a curious student arises naturally. In this definition, why does the co-domain of the multilinear mapping Φ have to be the real numbers $\mathbb{R}$? Can $\mathbb{R}$ be replaced by some other vector space? Is a multilinear mapping $\Psi : V \times \ldots \times V \to V$ a tensor? In particular, is a linear transformation $\varphi : V \to V$ a tensor?

The answer to these questions is that this definition is only a model of tensors. A cat is an example (model) of animals, while not all the animals are cats. There are other models of tensors which are not covered in this definition. We shall show (see more in Sec. 8 of Chap. 5) that indeed a multilinear mapping $\Psi : V \times \ldots \times V \to V$ is a vector-valued tensor. In particular, a linear transformation $\varphi : V \to V$ is a tensor. A quadratic form $\phi : V \to \mathbb{R}$ is also a tensor (quadratic forms are closely related to bilinear forms; see Appendix 1).

The following defines a tensor space (tensor product space). Then an element of this space is called a tensor. This is the abstract approach, and this is what we are going to adopt in the main course of this book (see Chap. 5).

Definition 5. (Tensor product space) Let U, V and W be vector spaces, and $\otimes : U \times V \to W$ be a bilinear mapping. The pair $(W, \otimes)$ is called a **tensor product space** (or simply **tensor space**) over the underlying vector spaces U and V, if they satisfy the following conditions:
(1) Generating property

$$W = \langle \mathrm{Im} \otimes \rangle ;$$

(2) Maximal span property

$$\dim W = \dim U \cdot \dim V.$$

The vectors in W are called **tensors** over U and V. The mapping $\otimes$ is called the **tensor multiplication** of two vectors, or **tensor product mapping**, or simply **tensor product**, or **tensor mapping**. W is often denoted by $U \otimes V$.

Remark. The coordinate change laws in the old-fashioned definition are only the phenomena. The essence of tensors is the multilinearity, or multilinear mappings. The coordinate change laws are the consequences of the multilinear mapping—tensor product mapping. In history, the multilinearity was understood by Gibbs and Ricci but was not emphasized explicitly.

The following definition is often seen in textbooks in pure mathematics.

Definition 6. Let U, V and W be vector spaces and suppose $\otimes : U \times V \to W$ is a bilinear mapping. $(W, \otimes)$ is called a **tensor product space** of U and V if the following conditions are satisfied (unique factorization property):

For any vector space X and any bilinear mapping $\Psi : U \times V \to X$, there exists a *unique* linear mapping $\varphi : W \to X$ such that

$$\Psi = \varphi \circ \otimes.$$

Remark. Some authors prefer this definition because it is terse in language, and it applies not only when U and V are finite dimensional spaces, but also when they are infinite dimensional vector spaces. It is not a good choice as a definition from the perspective of pedagogy for beginners. We shall treat this as a theorem about the universal property after the tensor product space is defined in an alternative way.

The following definition is based on construction (see the *Encyclopedic Dictionary of Mathematics* [Mathematical Society of Japan (1993)]; see also [Bourbaki (1942); Roman (2005)]). It describes the intuitive ideas of Gibbs dyadics but it is made rigorous in modern abstract language.

Definition 7. Let U and V be vector spaces over the same field F. Let $\mathscr{V}_F\langle U \times V\rangle$ be the free vector space generated by $U \times V$. Let Z be the subspace of $\mathscr{V}_F\langle U \times V\rangle$ generated by all the elements of the form

$$a(\mathbf{u}_1, \mathbf{v}) + b(\mathbf{u}_2, \mathbf{v}) - (a\mathbf{u}_1 + b\mathbf{u}_2, \mathbf{v}),$$
$$a(\mathbf{u}, \mathbf{v}_1) + b(\mathbf{u}, \mathbf{v}_2) - (\mathbf{u}, a\mathbf{v}_1 + b\mathbf{v}_2),$$

for all $a, b \in F$, $\mathbf{u}, \mathbf{u}_1, \mathbf{u}_2 \in U$ and $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$.

The quotient space

$$U \otimes V = \frac{\mathscr{V}_F\langle U \times V\rangle}{Z}$$

is called the **tensor product** of U and V. The elements in $U \otimes V$ are called **tensors** over U and V.

Define a mapping $\otimes : U \times V \to U \otimes V$ such that for all $\mathbf{u} \in U$ and $\mathbf{v} \in V$, $(\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} \otimes \mathbf{v} \overset{\text{def}}{=} [(\mathbf{u}, \mathbf{v})]$, where $[(\mathbf{u}, \mathbf{v})]$ is the equivalence class of $(\mathbf{u}, \mathbf{v})$ in $\mathscr{V}_F\langle U \times V\rangle$ defined by the subspace Z. This mapping is a bilinear mapping and is called the **canonical bilinear mapping**.

We have listed many different definitions of the tensor, which are commonly seen in textbooks. All of these are not exactly equivalent (some of them do, in some sense), but rather they reflect the historical evolution of the tensor concept.

* **Historical Note**: Evolution of definitions in mathematics

Many mathematical concepts are complex and difficult in nature. These concepts were not crystal clear when they were initially invented. These concepts have an evolutionary history and the definitions have been refined through time. Such examples are abundant, such as complex numbers, irrational numbers, real numbers, vectors, length, area, volume, probability, function, continuous function, Dirac delta function, infinity, infinitesimal, set, etc. Tensor is just one more example which can be added to the list. There have been occasions when a mathematician defined a new concept, it was even difficult for his contemporary fellow mathematicians to understand. Take Grassmann's exterior algebra for example. Heinrich Baltzer wrote to August Möbius after reading Grassmann's book *Ausdehnungslehre*: "It is not now possible for me to enter into those thoughts; I become dizzy and see sky-blue before my eyes when I read them." Möbius replied: "If as you write me, you have not relished Grassmann's *Ausdehnungslehre*, I reply that I have the same experience. I likewise have managed to get through no more than the first two sheets of his book."

* **Historical Note**: What are vectors exactly?

The concept of vector has gone through a similar long history of evolution as well. Some physical quantities like velocity and force are quantities with a magnitude and a direction. The parallelogram law of vector addition was known in Newton's time but the name vector was not used. The name vector was coined by Hamilton to denote the imaginary part $bi + cj + dk$ of his quaternion $a + bi + cj + dk$. It was Gibbs and Heaviside who liberated the vector from the shackles of the quaternion and made it an independent entity. At that time, vectors were mainly confined to three dimensions. This was soon generalized to higher dimensions and a vector was defined as an n-tuple. It was Peano who defined the vector space in the abstract sense in 1888. However, he did not use the name vector space, or linear space, but rather he called it a "linear system". (Interestingly, compare with the history of tensors. Ricci did not use the name "tensor", but rather a "system" instead.) Look at the following definitions of a vector.

(1) A vector is a quantity with a magnitude and a direction.

(2) A vector is an n-tuple of numbers.

(3) A vector is an element in a vector space.

These are not exactly equivalent definitions, but rather they reflect the historical evolution of the concept. Definition (2) is in terms of components. Definition (3) is abstract and axiomatic. With the definitions (2) and (3), a vector does not automatically have a magnitude.

A high school student often learns (1) as the definition of a vector in a physics course, but (2) as the definition in a mathematics course. He is likely to be confused with the question: are the vectors in physics and mathematics the same thing? The confusion shall be cleared when they learn the abstract definition of vector space in college, because (1) and (2) are just models of the abstract vectors.

The history of tensors is along a similar line. In this book, we are going to study the abstract, or axiomatic definition, and relate different concrete models to it.

* **Historical Note**: What are imaginary numbers exactly?

The typical definition of complex number in high school textbooks is: A complex number is *a number* that can be written in the form $a + bi$, where a and b are real numbers and i is the imaginary unit defined by $i^2 = -1$. This definition follows Jerome Cardan, who conceived it in 1545 without a solid logical foundation. The concept then kept evolving in the next three centuries to come, going through the initial confusion and denial to the final clarification and acceptance. Cardan himself considered these numbers as "mental tortures" and "useless". Descartes coined the term "imaginary" and rejected it. It was Gauss who named it "complex number" to rescue it from the mystery of the "imaginary" domain. Even Euler made a mistake in writing $\sqrt{-1}\sqrt{-4} = \sqrt{4} = 2$ in his book *Algebra*. It is a paradoxical argument by applying $\sqrt{a}\sqrt{b} = \sqrt{ab}$ to obtain $\sqrt{-1}\sqrt{-1} = \sqrt{(-1)(-1)} = 1$ (or similarly, $i^2 = \left(\sqrt{-1}\right)^2 = \sqrt{(-1)^2} = 1$).

The geometrical representation due to Argand marked a big step toward demystifying imaginary numbers. The modern definition of complex number is due to Hamilton in 1837: A complex number is an ordered pair (a, b) of real numbers. The number $(a, 0)$ is identified with the real number a, and i is defined as the pair $(0, 1)$. The addition and multiplication of complex numbers are defined by

$$(a_1, b_1) + (a_2, b_2) \overset{\text{def}}{=} (a_1 + a_2,\ b_1 + b_2),$$

$$(a_1, b_1) \cdot (a_2, b_2) \overset{\text{def}}{=} (a_1 a_2 - b_1 b_2,\ a_1 b_2 + a_2 b_1).$$

By this definition, $i^2 = (0, 1) \cdot (0, 1) = (-1, 0) = -1$.

* **Historical Note**: What are irrational numbers exactly?

This is basically the same question as "what are the real numbers exactly", because an irrational number can be defined as a real number that is not a rational number. Rational numbers are easier to define. The essence of a rational number is the ratio of two integers. A rational number can be defined as the equivalence class of a pair of integers. To many people's surprise, the concept of real numbers is much more complex than complex numbers. Logically, the concept of real numbers should precede that of complex numbers because a complex number is defined as a pair of real numbers, but historically, the rigorous definition of real numbers came much later than that of complex numbers. The concept of irrational numbers emerged from incommensurable segments in ancient Greek geometry and was used intensively in the early development of calculus without a rigorous definition. The rigorous definitions of real numbers, like Dedekind cuts and Cantor's construction through Cauchy sequences, finally came in the nineteenth century. In this sense, the complex number $\sqrt{-1}$ is much simpler than $\sqrt{2}$, because the latter involves infinite sets.

* **Historical Note**: What are sets exactly?

Georg Cantor was the founder of set theory, which serves as the foundation of modern mathematics. The concept of set, as a collection of objects, is intuitive. However, it is not precise. For example, we could think of a set U, which is the set of all sets. Since U is also a set, it is a member of itself—$U \in U$. There are other sets x with the property $x \notin x$. This leads to the Russell's paradox. Let us construct a set $Q \overset{\text{def}}{=} \{x | x \notin x\}$. Now we ask the question: is Q a member of itself?

Namely, is $Q \in Q$ true? First, suppose $Q \in Q$. Then Q does not satisfy the property $x \notin x$, and hence $Q \notin Q$. Next, suppose $Q \notin Q$. Then Q satisfies the property $x \notin x$. Hence $Q \in Q$. A popular version of this is the barber paradox: a barber in a village, who is a man, claims that he shaves every man in the village who does not shave himself, and does not shave any man who shaves himself. Now there is a question: does the barber shave himself? According to his claim, he shaves himself if and only if he does not shave himself.

Gottlob Frege was a German logician, who made significant contributions in logic. Russell's paradox was a big blow to him. He became depressed and did no serious mathematics thereafter. Unlike physicists (see Sec. 6 of Chap. 10; see also [Guo (2021)][3]), mathematicians take paradoxes seriously. What is a way out of this paradox? It is actually pretty simple. We redefine the concept of set more precisely so that those trouble makers like U and Q no longer qualify to be called sets. It is not an ordinary definition. The qualification is regulated by a set of axioms introduced by Zermelo and Fraenkel. These axioms are actually the hidden definition of set (see more on axiomatic systems in Appendix 3).

§4. Plain Things by Fancy Tensor Names

Quite some terms bear the surname "Tensor", like metric tensor, curvature tensor, inertia tensor, stress tensor, diffusion tensor imaging, etc. These are just fancy names for plain things, which may sound intimidating to beginners. Yes, they are tensors and it is not wrong to call them tensors, but tensor theory is not essential to understand these concepts. They can go by other names without the use of "tensor". Calling them tensors is like calling water by the name "dihydrogen monoxide". Everyone understands water, but people may be confused by the chemistry jargon.

These terms were named historically because of the fact that they are (represented by) matrices. The confusion is rooted in the question whether a tensor is the same as a matrix. If it does, why don't we simply call them metric matrix, inertia matrix, etc.? The old-fashioned definition of tensor is equivocal about whether a tensor is simply a matrix or not. A tensor is defined as a matrix of components, but amended awkwardly by the transformation laws.

[3]Guo, H. (2021). A New Paradox and the Reconciliation of Lorentz and Galilean Transformations, *Synthese*, https://doi.org/10.1007/s11229-021-03155-y (open access).

Things get clear with the modern view. The metric tensor is just an inner product, the inertia tensor can be defined as a linear transformation or a quadratic form. The stress tensor and diffusion tensor are simply linear transformations. We shall discuss inertial tensor in more detail in Chap. 8, and the metric tensor for Riemannian geometry in Chap. 10.

Think of the stress forces in liquids and solids. In a liquid, let us single out a small piece of imaginary surface, which separates the liquid on both sides. Each side exerts a force on the other side (Figure 1.1a). Let us use a vector **S** to represent the surface, where **S** is a normal vector of the surface, and the magnitude of **S** represents the area of the surface. Let **F** be the vector representing the force that the liquid on one side exerts on the other side. Because liquids cannot have shear forces, the force **F** must be in the normal direction of the surface, which is the same as **S**. **F** is linearly related to **S**,

$$\mathbf{F} = \sigma \mathbf{S}, \tag{1.11}$$

where σ is a scalar coefficient, which is called the pressure.

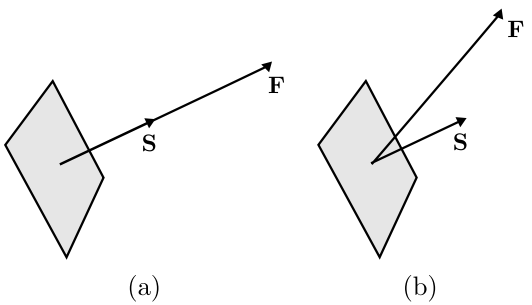

Figure 1.1 (a) Stress in liquids (b) Stress in solids

Things are different in solids, like crystals. The force **F** in general is not in the same direction as **S**. **F** can be decomposed into normal stress, and shear stress (in the tangent direction of the surface). However, **F** is still linearly related to **S** (Figure 1.1b). This relation is a linear transformation:

$$\mathbf{F} = \Sigma \mathbf{S}, \tag{1.12}$$

where Σ is a linear transformation which can be represented by a matrix $[\Sigma]$ with components σ_{ij},

$$\begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix}.$$

Σ is called the stress tensor. This can be written as

$$F_i = \sum_{j=1}^{3} \sigma_{ij} S_j. \tag{1.13}$$

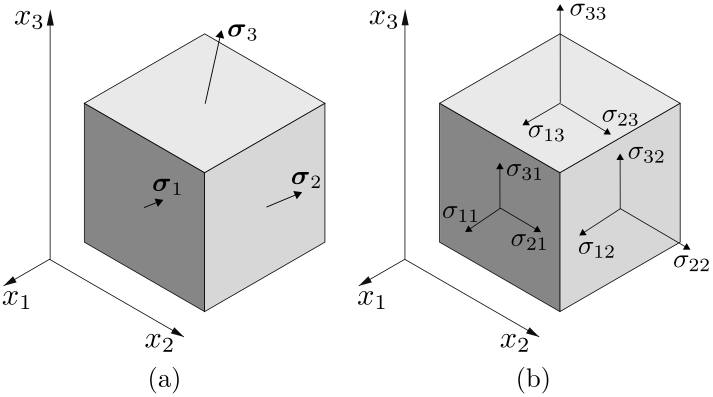

Figure 1.2 (a) Stress tensor as three vectors (b) The nine components of the stress tensor

The matrix of the stress tensor Σ can be viewed as three column vectors

$$\boldsymbol{\sigma}_1 = \begin{bmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{31} \end{bmatrix}, \boldsymbol{\sigma}_2 = \begin{bmatrix} \sigma_{12} \\ \sigma_{22} \\ \sigma_{32} \end{bmatrix}, \boldsymbol{\sigma}_3 = \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \\ \sigma_{33} \end{bmatrix}.$$

What are the physical meanings of these three vectors? Imagine we have a small cube. Their faces are along the three axes with normal vectors $\mathbf{s}_1 = (1, 0, 0)$, $\mathbf{s}_2 = (0, 1, 0)$, $\mathbf{s}_3 = (0, 0, 1)$ and unit area. $\boldsymbol{\sigma}_1$ is the stress force acted on the face $\mathbf{s}_1$, $\boldsymbol{\sigma}_2$ is the stress force acted on the face $\mathbf{s}_2$, and so on (Figure 1.2a). Each force $\boldsymbol{\sigma}_i$ has three components and together the stress matrix has nine components. What is the physical meaning of the component σ_{ij}? σ_{ij} represents the ith component of $\boldsymbol{\sigma}_j$, which is the force acting on the face $\mathbf{s}_j$ (orthogonal to x_j axis). On face $\mathbf{s}_1$, σ_{11} is the normal stress while σ_{21} and σ_{31} are the tangent stresses. On face $\mathbf{s}_2$, σ_{22} is the normal stress while σ_{12} and σ_{32} are the tangent stresses (Figure 1.2).

In fact, the tensor here is just a linear transformation, and the stress tensor Σ is just one example of linear transformations used in physics. Eq. 1.13 is the component form of any linear transformation, not just limited to the stress situation. The linear transformation maps any vector $\mathbf{S}$ to a new vector $\mathbf{F} = \Sigma\mathbf{S}$, as in Eq. 1.12. The meaning of its component σ_{ij} is the ith component of $\mathbf{F}$ when $\mathbf{S}$ is a unit vector along the jth direction. Here we have given a physical interpretation of the linear transformation Σ in the example of stress in solids, or crystals.

The physical process of diffusion in isotropic media is described by Fick's law:

$$\mathbf{J} = -d\nabla\phi,$$

where ϕ is the concentration density of the diffusive substance, which is a function of the spatial location $\mathbf{x}$; $\nabla\phi$ is the gradient of ϕ; $\mathbf{J}$ is the flux of the diffusive substance, and d is a scalar constant called the diffusion coefficient. However, in anisotropic media, the flux $\mathbf{J}$ is usually not in the same direction as $\nabla\phi$, but it still has a linear relationship with $\nabla\phi$. This means that $\mathbf{J}$ and $\nabla\phi$ are related by a linear transformation:

$$\mathbf{J} = -D\nabla\phi.$$

This linear transformation D is often called the diffusion tensor and it has nine components when a coordinate system is chosen. In coordinate form, it can be written as

$$J_i = -\sum_{j=1}^{3} D_{ij}\frac{\partial\phi}{\partial x_j}.$$

The brain consists of gray matter and white matter. The gray matter consists of the neuron bodies while the white matter consists of the myelinated axon fibers, which serve as the interconnections between the neurons. The diffusion of water in the brain is highly anisotropic due to these axon fibers. With the help of magnetic resonance imaging (MRI), the diffusion tensor components at space locations can be measured, which is used to reconstruct the fiber tracts in the brain. This is known as diffusion tensor imaging (DTI). Figure 1.3 shows the diffusion tensor field (represented by ellipsoids, see Sec. 5 of Chap. 8). Figure 1.4 shows the reconstructed fiber tracts of the brain using DTI.

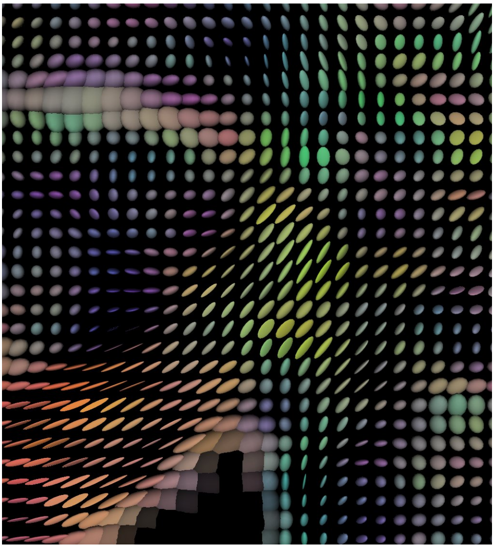

Figure 1.3 Diffusion Tensor Imaging: ellipsoids of the diffusion tensors

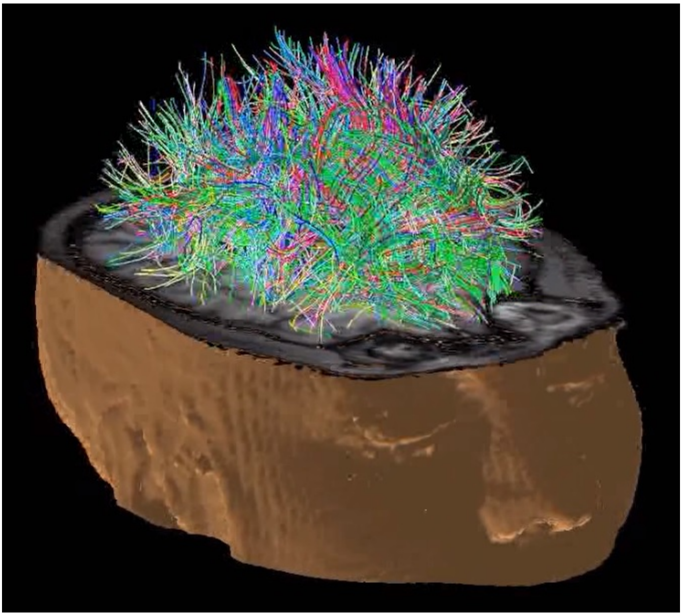

Figure 1.4 Diffusion Tensor Imaging: fiber tracks in the brain white matter

§5. Tensors without a Tensor Name— Linear Transformations

Many objects that we are familiar with are actually tensors, but they do not often go by a tensor name. We shall show that linear mappings and linear transformations are tensors. Realizing these mundane objects are actually tensors has a demystifying effect. Here is just the gospel. The details will be discussed in Chaps. 5 and 6.

When a basis of the vector space V is chosen, a linear transformation $\varphi : V \to V$ can be represented by a matrix. When the basis is changed, the matrix of the linear transformation changes in accordance. This explains why the tensors in the old-fashioned definition have to obey the transformation laws, and most importantly, it explains what causes the transformations.

Suppose $\langle \cdot, \cdot \rangle$ is an inner product defined in V. Given two constant vectors $\mathbf{a}, \mathbf{b} \in V$, we define a linear transformation:

$$\varphi_{\mathbf{a},\mathbf{b}} : V \to V;$$
$$\mathbf{x} \mapsto \varphi_{\mathbf{a},\mathbf{b}}(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{a} \langle \mathbf{b}, \mathbf{x} \rangle\,, \text{ for all } \mathbf{x} \in V.$$

Basically, the vector $\mathbf{x}$ is projected onto $\mathbf{b}$ and the inner product $\langle \mathbf{b}, \mathbf{x} \rangle$ is calculated. The final output is a vector along the direction of $\mathbf{a}$ but scaled by the factor $\langle \mathbf{b}, \mathbf{x} \rangle$.

The vector $\mathbf{b}$ here can be viewed as a linear function in the dual space V^*. The effect of $\mathbf{b}$ acting on a vector $\mathbf{x} \in V$ is $\mathbf{b}(\mathbf{x}) = \langle \mathbf{b}, \mathbf{x} \rangle$. The linear transformation $\varphi_{\mathbf{a},\mathbf{b}}$ is actually the tensor product in $V \otimes V^*$ and we denote $\varphi_{\mathbf{a},\mathbf{b}} = \mathbf{a} \otimes \mathbf{b}$.

A beginner might be tempted to guess that all the linear transformations can be put in the form of $\mathbf{a} \otimes \mathbf{b}$, for some $\mathbf{a} \in V$ and $\mathbf{b} \in V^*$, but this is not true. However, any linear transformation can be written as the sum of these tensor products, $\mathbf{a}_1 \otimes \mathbf{b}_1 + \ldots + \mathbf{a}_k \otimes \mathbf{b}_k$. Therefore, a linear transformation is a mixed tensor of type $(1, 1)$, and of course, it obeys the transformation law in Eq. 1.4. This is also why the inertia tensor, stress tensor and diffusion tensor are tensors, but using plain words, they are just linear transformations.

A linear transformation is also a special case of a more general model—vector-valued tensor, which is a multilinear mapping $\Phi : V_1 \times \ldots \times V_q \to X$. When $q = 1$ and $V_1 = X = V$, we have a linear transformation $\Phi : V \to V$. We discuss vector-valued tensors in Sec. 8 of Chap. 5.

§6. Comparison: Different Definitions of the Vector —Concrete Systems vs. Abstract Systems

To better understand the concept of tensor, we make a comparison with the vector, which we are already familiar with. The key to understand the difficulty associated with tensors is the appreciation of the relationship between the abstract concepts and concrete examples.

Historically, there have been different definitions of vectors too. These definitions are not exactly equivalent and they reflect the historical evolution of the concept.

Definition 8. A vector is a quantity with a magnitude and a direction.

Definition 9. A vector is a directed line segment in space. The addition of two vectors is defined by the parallelogram law.

Definition 10. A vector is an n-tuple of real numbers $(x_1, \ldots, x_n)$.

Definition 11. Let F be a field and V a nonempty set. V together with two operations called addition $(+) : V \times V \to V$ and scalar-vector multiplication $() : F \times V \to V$, is called a vector space over F, if these operations satisfy the following conditions. The elements in V are called vectors and the elements of F are called scalars.

(1) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
(2) There exists $\mathbf{0} \in V$ such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.
(3) For any $\mathbf{u} \in V$, there exists $\mathbf{x} \in V$ such that $\mathbf{u} + \mathbf{x} = \mathbf{0}$. We denote $\mathbf{x} = -\mathbf{u}$.
(4) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.
(5) $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$.
(6) $a(b\mathbf{u}) = (ab)\mathbf{u}$.
(7) $1\mathbf{u} = \mathbf{u}$, where $1 \in F$ is the multiplicative identity in F.

A reader may have already learned that the vector space is an Abelian (commutative) group with respect to the vector addition, but finds that the commutative law $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ is missing from the above list of axioms. These axioms were first proposed by Peano. He included this commutative law and almost all the textbooks afterwards just followed him. However, this axiom is not independent of the rest, and hence there is no need to list it explicitly (see a proof in Appendix 1). Peano was a master with the axiomatic systems. It is remarkable that he devised this axiomatic system for vector space (which he called linear system) as early as 1888. Amazingly all of the axioms, except the commutative law of addition, turned out to be independent.

Remark. Definition 8 is traditional and vague. Definition 10 is more general than Definition 9, as it defines an n-dimensional vector while the vector in Definition 9 is 3-dimensional.

Definition 11 is the most general and the most abstract of all. It is an axiomatic definition. Any system that satisfies these axioms is called a model of the abstract vector space. Vectors defined in Definitions 9 and 10 are examples, or models of a vector space. We can find many other models of vectors in the following.

Example 1. (Matrix spaces) All $m \times n$ real matrices $M_{m,n}$ form a real vector space with respect to matrix addition and matrix multiplication by a number. Each $m \times n$ matrix is a vector.

Example 2. (Linear mappings) Let V and W be vector spaces. All linear mappings $\varphi : V \to W$ form a vector space. Each linear mapping is a vector.

Example 3. (Polynomials of degree at most n) All polynomials with real coefficients of degree at most n, form a real vector space with respect to polynomial addition and multiplication by a number. Each polynomial is a vector.

Example 4. (All polynomials) All polynomials of one variable with real coefficients form a real vector space with respect to addition and multiplication by a number. Each polynomial is a vector. This vector space is infinite dimensional.

Example 5. (Real functions) All real functions $f : \mathbb{R} \to \mathbb{R}$ form a real vector space. If f, g are two real functions and $a, x \in \mathbb{R}$, we define $f+g = h$, where $h(x) = f(x) + g(x)$; and $(af)(x) = af(x)$. Each real function is a vector. This vector space is infinite dimensional.

Despite the large number of apparently different models, there is one interesting property. That is, any model of an n-dimensional vector space is isomorphic to each other, in particular, isomorphic to the vector space of

n-tuples in Definition 10. Because of this isomorphism, we have the liberty of choosing the abstract Definition 11, or the concrete Definition 10.

The different definitions of tensors also reflect the history of evolution of the concept.

Definition 5 for tensors is in a similar position to Definition 11 for vectors. It is an abstract or axiomatic definition. Definitions 3, 4 and 7 are models of the abstract tensor.

§7. Tensor Product and Tensor Spaces

We can ask two different but related questions:

"What is a tensor?"

"What is a tensor space?"

Definitions 3 and 4 define an individual tensor, while Definition 5 defines an abstract tensor (product) space $U \otimes V$, and any element in this space is called a tensor.

We shall discuss tensor product spaces in Chap. 5 and tensor power spaces $V^{\otimes p} = V \otimes \ldots \otimes V$ in Chap. 6.

When we talk about tensor spaces $U \otimes V$ or $V^{\otimes p}$, we should not neglect the relationship between the tensor space $V^{\otimes p}$ and the vector space V. We call V the *underlying vector space* of tensor space $V^{\otimes p}$.

There is a good comparison with vector spaces. Recall, in a vector space, there are two distinct sets, the set of vectors V and the set of scalars, which is a field F. V is called the "vector space *over* the field F" and F is called the ground field of V (Figure 1.5).

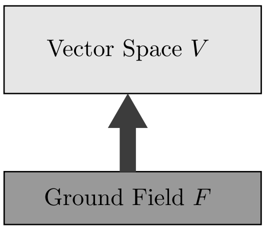

Figure 1.5 Vector space V and its ground field F

The interaction between the ground field F and vector space V is through the scalar-vector multiplication $() : F \times V \to V$.

The relationship between the underlying vector space and the tensor space is the tensor product, which is a bilinear mapping $\otimes : V \times V \to V \otimes V$. From this point of view, the tensor space $V^{\otimes 2} = V \otimes V$ is a vector space by itself. A tensor is also a vector. This view is different from the traditional view that tensors are generalizations of vectors because their transformation laws are different (Figure 1.6).

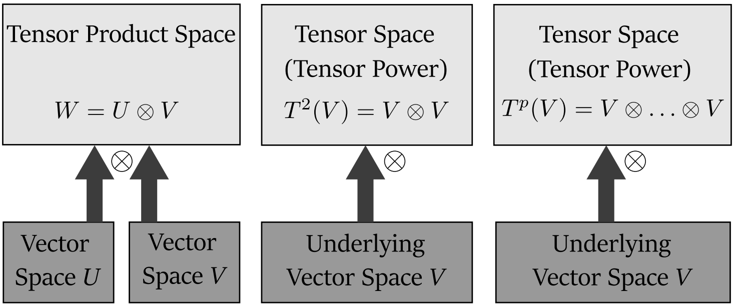

Figure 1.6 Tensor space $V^{\otimes p}$ and its underlying vector space V

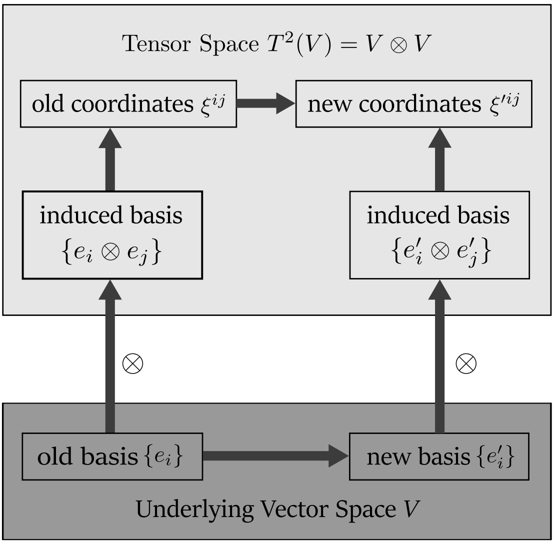

Figure 1.7 Coordinate change of a tensor

Given a basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for V, the tensors $\{\boldsymbol{\tau}_{ij} | \boldsymbol{\tau}_{ij} = \mathbf{e}_i \otimes \mathbf{e}_j,\ i, j = 1, \ldots, n\}$ form a basis for the tensor space $V^{\otimes 2}$, which contains n^2 basis vectors. When the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of V changes, the induced basis $\{\boldsymbol{\tau}_{ij}\}$ for $V^{\otimes 2}$ changes to $\{\boldsymbol{\tau}'_{ij}\}$ accordingly. Then the change of coordinates of a tensor in $V^{\otimes 2}$ obeys those laws in Definition 2. Therefore those coordinate change laws refer to coordinate changes of the tensors in $V^{\otimes 2}$ in response to the basis change of V, rather than in response to the basis change of $V^{\otimes 2}$ itself, which is also a vector space (Figure 1.7). Therefore, a tensor is also a vector, rather than a generalization of a vector. We could use a single index running from 1 to n^2 for the tensor components. If its basis changes, the components of a tensor in $V^{\otimes 2}$ with a single index will just behave like a vector (Figure 1.8). The reason we adopt double indices ij is the relationship between V and $V^{\otimes 2}$, which is the tensor product $\otimes$.

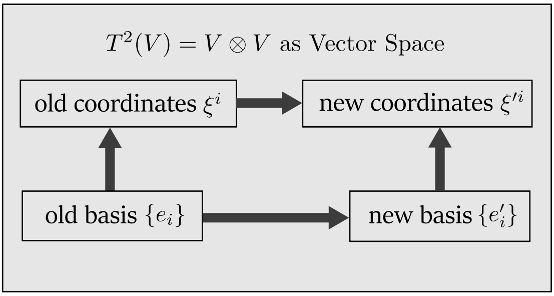

Figure 1.8 Coordinate change of a tensor as a vector

§8. Degree, Rank, Order or Dimension—Which Is the Best Name?

One may encounter a mixture of terms in literature—rank, order and degree, used interchangeably. They all mean the same thing, the number of indices of a tensor component. In the machine learning community, they even use "dimension" for this, because they use the term tensor as a multi-dimensional array.

Ricci never used the term "tensor" in his writings. He called it a "system". He also used the term "order" of a system. Physicists use "rank" more often.

However, in the modern view, the tensor space $V^{\otimes p} = V \otimes \ldots \otimes V$

is the p-th tensor power (tensor product of the same vector space with itself p times). It is natural to call p the degree, drawing similarity with the naming of the degree of polynomials. This naming agrees with the *Encyclopedic Dictionary of Mathematics* [Japanese Mathematical Society (1993)], which is an excellent reference source and provides the standard terminology of modern mathematics.

Following N. Bourbaki [(1942)], the term rank of a tensor is defined with a different meaning from the degree. Recall the rank of a square matrix (similarly for a linear transformation) is defined as the number of linearly independent columns (or rows). An $n \times n$ square matrix can have any rank between 1 and n. A tensor of degree 2 may have any rank between 1 and n. Any decomposable tensor of degree 2 has a rank of 1 (see more in Sec. 5 of Chap. 5).

*§9. What Are Pseudo-Scalars, Pseudo-Vectors and Pseudo-Tensors Exactly?

In older physics textbooks, some authors introduce the concepts of pseudo-scalars, pseudo-vectors, and in general pseudo-tensors. They are also defined by different transformation laws. Let us first look at the so called pseudo-vectors.

This is the definition: a quantity is called a pseudo-vector (or axial vector) if it transforms like a vector under proper transformation (for example, rotation), but the transformation gains an additional sign flip under an improper transformation.

A proper transformation reserves the orientation of an oriented vector space while an improper transformation changes the orientation. For example, the reflection $x' = -x$, $y' = -y$, $z' = -z$ is an improper transformation.

One example of a pseudo-vector is illustrated as the cross product $\mathbf{w} = \mathbf{u} \times \mathbf{v}$. They argue, for a regular vector (also called polar vector), when the coordinates go through a reflection, $\mathbf{v}$ should be transformed to $\mathbf{v}' = -\mathbf{v}$. But for the cross product, $\mathbf{w}' = (-\mathbf{u}) \times (-\mathbf{v}) = \mathbf{w}$. Magnetic field and angular momentum are examples of pseudo-vectors.

One example of a pseudo-scalar is the triple scalar product (representing signed volume) of three vectors $a = \mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3)$. When the coordinates go through a reflection, $a' = (-\mathbf{v}_1) \cdot [(-\mathbf{v}_2) \times (-\mathbf{v}_3)] = -a$.

This argument does not seem to make sense. A scalar is just a number and it should not depend on coordinates. Why should it be affected by coordinate reflection and change sign accordingly?

A closer examination reveals that something is not expressed clearly and logically in these concepts. We take the pseudo-vector for example. Let V and W be 3-dimensional vector spaces, $\mathbf{u}, \mathbf{v} \in V$ and $\mathbf{w} \in W$. V and W are isomorphic, but let us distinguish them. Now we view the cross product as a mapping $(\times) : V \times V \to W$. Here $\times$ is not a tensor product mapping, but it is a bilinear mapping in a similar situation. It connects spaces V and W. Let $\mathbf{w} = \mathbf{u} \times \mathbf{v}$, and let $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ be a basis for V. We define $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in W$,

$$
\begin{aligned}
\mathbf{e}_1 &\stackrel{\text{def}}{=} \mathbf{b}_2 \times \mathbf{b}_3, \\
\mathbf{e}_2 &\stackrel{\text{def}}{=} \mathbf{b}_3 \times \mathbf{b}_1, \\
\mathbf{e}_3 &\stackrel{\text{def}}{=} \mathbf{b}_1 \times \mathbf{b}_2.
\end{aligned} \tag{1.14}
$$

Then $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ forms a basis for W. After coordinate reflection, the new induced basis vectors are

$$
\begin{aligned}
\mathbf{e}'_1 &\stackrel{\text{def}}{=} \mathbf{b}'_2 \times \mathbf{b}'_3 = (-\mathbf{b}_2) \times (-\mathbf{b}_3) = \mathbf{e}_1, \\
\mathbf{e}'_2 &\stackrel{\text{def}}{=} \mathbf{b}'_3 \times \mathbf{b}'_1 = (-\mathbf{b}_3) \times (-\mathbf{b}_1) = \mathbf{e}_2, \\
\mathbf{e}'_3 &\stackrel{\text{def}}{=} \mathbf{b}'_1 \times \mathbf{b}'_2 = (-\mathbf{b}_1) \times (-\mathbf{b}_1) = \mathbf{e}_3.
\end{aligned}
$$

Therefore, $\mathbf{w}$ has the same coordinates under induced basis $\{\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3\}$ as under basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. This is also explained with Figure 1.7 in a similar way, except now the mapping is the cross product $\times$, instead of the tensor product $\otimes$. This means, as a 3-tuple and a member of W, $\mathbf{w}$ is certainly an ordinary vector. If the space W is unrelated to V, when the basis of W goes through a reflection, the coordinates of $\mathbf{w}$ with respect to the new basis of W certainly flip the sign. When we say $\mathbf{w}$ is a pseudo-vector and the signs of $\mathbf{w}$ do not change, we are talking with respect to the induced basis $\mathbf{e}'_1 = \mathbf{b}'_2 \times \mathbf{b}'_3$, $\mathbf{e}'_2 = \mathbf{b}'_3 \times \mathbf{b}'_1$ and $\mathbf{e}'_3 = \mathbf{b}'_1 \times \mathbf{b}'_2$, which are induced by the cross product.

After all, the pseudo-vectors can be viewed as living in a vector space W. The pseudo-vectors are just ordinary vectors and transform as ordinary vectors with respect to a basis change in W itself. However, there is a connection between the vector space W with another underlying vector space V. In general, let us denote it by $\circledast : V \times V \to W$. The coordinates of a pseudo-vector in W changes like a pseudo-vector with respect to basis change in V composed with the mapping $\circledast$.

The cross product only applies in 3-dimensional vector spaces. For the general n-dimensional vector space V, the pseudo-vectors can be viewed as living in the space of $\Lambda^{n-1}(V)$, which is the exterior space over V to the $(n-1)$-th power. It has the same dimension as V. The pseudo-vector in $\Lambda^{n-1}(V)$ can be viewed as the Hodge dual of a vector in V. A pseudo-scalars can be viewed as living in the space of $\Lambda^n(V)$, which is the dual of $\Lambda^0(V) \stackrel{\text{def}}{=} \mathbb{R}$ and has dimension 1.

For a pseudo-tensor of degree two, it transforms as

$$(\xi')^{st} = \operatorname{sign}(\Lambda) \sum_{\sigma,\tau} \xi^{\sigma\tau} \Lambda_\sigma{}^s \Lambda_\tau{}^t,$$

where $\operatorname{sign}(\Lambda)$ is the sign of $\det \Lambda$. This extra sign can also be viewed as the result of some bilinear mapping connecting the space of pseudo-tensors W to the underlying vector space V,

$$\circledast : V \times V \to W.$$

The more general concept is the tensor density of weight k, with a transformation law

$$(\xi')^{st} = (\det \Lambda)^k \sum_{\sigma,\tau} \xi^{\sigma\tau} \Lambda_\sigma{}^s \Lambda_\tau{}^t,$$

where $\det \Lambda$ is the determinant of the transformation matrix Λ in the underlying vector space V, and k is a constant exponent.

§10. What Is Tensor Analysis Exactly? Relation to Riemannian Geometry

10.1 Vector Analysis

Vector analysis studies vector-valued functions. Let V be a vector space over $\mathbb{R}$. A vector-valued function can be a function of a single variable $\mathbf{p} : \mathbb{R} \to V; t \mapsto \mathbf{p}(t)$, or a function of multiple variables, like $\mathbf{f} : \mathbb{R}^3 \to V; (x, y, z) \mapsto \mathbf{f}(x, y, z)$. $\mathbf{p}(t)$ is often interpreted as a vector which changes with time t, while $\mathbf{f}(x, y, z)$ is a vector field, with a vector $\mathbf{f}$ assigned to each spatial location (x, y, z). So vector analysis is the differential calculus of vector fields, while the single variable vector functions can be viewed as a special case.

Gibbs was a pioneer of vector analysis. His book [Gibbs (1884)] deals with both vector algebra and vector analysis. In vector analysis, three differential operators on vector (or scalar) fields are defined: the gradient of a scalar field $\nabla\varphi$, the divergence of a vector field $\nabla \cdot \mathbf{f}$ and the curl (or rot, for rotation) of a vector field $\nabla \times \mathbf{f}$. Important theorems involving these operators include Gauss' theorem

$$\iiint_V (\nabla \cdot \mathbf{f}) dV = \oiint_{\partial V} \mathbf{f} \cdot d\mathbf{S},$$

Stoke's theorem

$$\iint_S (\nabla \times \mathbf{f}) \cdot d\mathbf{S} = \oint_{\partial S} \mathbf{f} \cdot d\mathbf{r},$$

and properties like

$$\nabla \times (\nabla\varphi) = 0,$$
$$\nabla \cdot (\nabla \times \mathbf{f}) = 0.$$

10.2 Tensor Analysis and Riemannian Geometry

Some people view tensors as the generalization of vectors, and it is natural to guess that the study of tensors should be divided into tensor algebra and tensor analysis, with the latter studying the differential calculus of tensor fields in Euclidean space $\mathbb{R}^3$. As a matter of fact, tensor analysis in this sense was also developed by Gibbs in his book of vector analysis. Gibbs used different terminology but his dyadics and polyadics are just tensors in the modern sense. He defined several algebraic operations—dot products and cross products for dyads, which can be linearly extended to general tensors, like

$$\mathbf{a} \cdot (\mathbf{bc}) \stackrel{\text{def}}{=} (\mathbf{a} \cdot \mathbf{b})\mathbf{c},$$
$$(\mathbf{ab}) \cdot \mathbf{c} \stackrel{\text{def}}{=} \mathbf{a}(\mathbf{b} \cdot \mathbf{c})$$
$$(\mathbf{ab}) \cdot (\mathbf{cd}) \stackrel{\text{def}}{=} (\mathbf{b} \cdot \mathbf{c})\mathbf{ad},$$
$$\mathbf{a} \times (\mathbf{bc}) \stackrel{\text{def}}{=} (\mathbf{a} \times \mathbf{b})\mathbf{c},$$
$$(\mathbf{ab}) \times \mathbf{c} \stackrel{\text{def}}{=} \mathbf{a}(\mathbf{b} \times \mathbf{c}),$$
$$(\mathbf{ab}) : (\mathbf{cd}) \stackrel{\text{def}}{=} (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}),$$

etc. Along this line, viewing the nabla operator ∇ as a vector operator, the gradient of a vector $\nabla\mathbf{u}$, the gradient, divergence and curl of tensors $\nabla(\mathbf{uv})$, $\nabla \cdot (\mathbf{uv})$, $\nabla \times (\mathbf{uv})$ and many other operations can be defined. Gibbs did explore the properties of these operations and demonstrated many applications in physics and mathematics, including applications to the curvature of surfaces in differential geometry.

However, tensor analysis in this direction of studying tensor fields in Euclidean space $\mathbb{R}^3$ has not gone too far in history, because it is kind of trivial. What is called tensor analysis today is in the context of Riemannian geometry. The tensor fields are assumed to be tensor fields on a Riemannian manifold, or a differentiable manifold in general.

Ricci called his work absolute calculus, with an emphasis on the covariant derivative. Levi-Civita contributed the concept of parallel transport. Levi-Civita did not use the term tensor in his early works, but adopted this new name in his book [Levi-Civita (1927)] *The Absolute Differential Calculus* (*Calculus of Tensors*) after Einstein and Grossmann had popularized the term tensor.

However, tensor analysis is not really a new branch, or independent branch of mathematics. It is just Riemannian geometry in a slightly different dialect, characterized by the component (or index) form of representation. In his *Mathematical Thought from Ancient to Modern Times*, M. Kline [(1972)] writes:

> "Tensor analysis is often described as a totally new branch of mathematics, created *ab initio* either to meet some specific objective or just to delight mathematicians. It is actually no more than a variant on an old theme, namely, the study of differential invariants associated primarily with a Riemannian geometry."

The "differential invariant associated primarily with a Riemannian geometry" that Kline refers to is the fundamental form $ds^2 = \sum_{i=1}^{n} g_{ij} dx_i dx_j$, or the line element, or the metric tensor, which is the higher dimensional generalization of Gauss' first fundamental form. It is invariant under coordinate transformations (or isometric mappings, in the active view), or re-parameterizations (the passive view). The characteristic of Ricci's absolute differential calculus, or tensor analysis is the component approach. É. Cartan [(2002)] recommended, "as far as possible avoid very formal computations in which an orgy of tensor indices hides a geometric picture which is often very simple." Chap. 10 provides an outlook of Riemannian geometry and general relativity but it is not the scope of this book to go deeper than that. The reader is referred to [Bishop and Goldberg (1980)] and [Guo (2014)] for further reading.