

A Mathematical construction of Contravariant/Covariant Vectors and Tensors for Physics Learners

Shunta Takahashi^{a,b}

^a*Department of Physics, Kyoto University, Kyoto 606-8502, Japan*

^b*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*

Contents

1	Introduction	1
2	Prerequisite knowledge: Einstein’s summation convention	2
3	Definition of contravariant and covariant vectors	2
3.1	Contravariant vectors	2
3.2	The dual vector space	3
3.3	Covariant vectors	3
4	The meaning of “contravariant” and “covariant”	4
4.1	Transformation of contravariant vector components	4
4.2	Transformation of covariant vector components	5
5	Tensors	6
5.1	Definition of a tensor	6
5.2	Tensor components and their transformation	7
6	Tensors as multilinear maps	8
7	Application in physics: the Minkowski metric	8
7.1	The Minkowski metric as a $(0, 2)$ -tensor	9
7.2	Raising and lowering indices	9
8	Conclusion	10

Contents

1 Introduction

Physics textbooks often introduce contravariant vectors, covariant vectors, and tensors by defining them through their component transformation rules under a change of coordinates. For many students, this approach can be indirect and complex to grasp intuitively. This article aims to provide a clearer path to understanding these concepts by starting with simpler, more fundamental mathematical definitions. By first defining contravariant and covariant vectors based on the algebraic structures of a vector space and its dual, their characteristic transformation properties emerge as a natural consequence. This foundation is then extended to develop a rigorous and intuitive understanding of tensors.

The discussion assumes a finite n -dimensional real vector space V with a basis $(e_i)_{1 \leq i \leq n}$. All concepts presented here can be generalized to complex vector spaces by simply replacing the field of real numbers \mathbb{R} with the field of complex numbers \mathbb{C} .

2 Prerequisite knowledge: Einstein's summation convention

Throughout this article, we will employ *Einstein's summation convention* to simplify notation. This convention dictates that whenever an index variable appears twice in a single term, once as an upper (superscript) index and once as a lower (subscript) index, summation is implied over all possible values of that index. The explicit summation symbol \sum is omitted.

For example, consider two vectors v and w in an n -dimensional space. Their dot product, typically written as $v \cdot w = \sum_{i=1}^n v_i w_i$, can be expressed by representing the components of v with an upper index $(v^i)_{1 \leq i \leq n}$ and the components of w with a lower index $(w_i)_{1 \leq i \leq n}$

$$v \cdot w = v^i w_i. \quad (2.1)$$

This compact notation implicitly represents the sum $v^1 w_1 + v^2 w_2 + \dots + v^n w_n$. Similarly, the product of two matrices A and B can be written using this convention. The element in the i -th row and j -th column of the product matrix $C = AB$ is given by $C_j^i = \sum_{k=1}^n A_k^i B_j^k$. Using the summation convention, this becomes

$$(AB)_j^i = A_k^i B_j^k. \quad (2.2)$$

The repeated index k (one upper, one lower) implies summation.

3 Definition of contravariant and covariant vectors

We begin by providing coordinate-free, mathematical definitions for contravariant and covariant vectors.

3.1 Contravariant vectors

The mathematical definition of a contravariant vector is remarkably straightforward.

Definition 3.1: Contravariant vector

A **contravariant vector** is an element of a vector space V .

That's it. A contravariant vector is simply an element of any vector space. For instance, if we consider the familiar vector space $V = \mathbb{R}^n$, then any column vector $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ is a contravariant vector. We will explain step-by-step how this connects to the physics-style definition. At this stage, it may not be clear why the term "contravariant" is used. This nomenclature arises from how the *components* of the vector transform under a change of basis, a topic we will explore in Section 4.

3.2 The dual vector space

To define covariant vectors, we must first introduce the concept of the dual vector space.

Definition 3.2:

Let V be a vector space over the field \mathbb{R} . The **dual vector space** of V , denoted V^* , is the set of all linear maps (also known as **linear functionals**) from V to \mathbb{R} .

$$V^* := \{T : V \rightarrow \mathbb{R} \mid T \text{ is a linear map}\} \quad (3.1)$$

The elements of V^* can be added and multiplied by scalars:

$$\begin{aligned} (T + S)(v) &:= T(v) + S(v), \\ (aT)(v) &:= a \cdot T(v) \quad (a \in \mathbb{R}). \end{aligned} \quad (3.2)$$

It is easy to see that V^* is in fact a vector space, but it can actually be confirmed as follows that if $\dim V = n$, then $\dim V^* = n$.

First, note that a linear map $T : V \rightarrow \mathbb{R}$ is determined solely by its values on the basis $(e_i)_{1 \leq i \leq n}$ of V . Indeed, for any $v = v^i e_i \in V$, the value of $T(v)$ is determined by the linearity of T as

$$T(v) = v^i T(e_i). \quad (3.3)$$

While a map between general sets must be defined by specifying a value for each individual argument, this special property occurs for linear maps. (In general, an R -homomorphism from a free module over a ring R to any R -module is determined only by its values on the basis.)

Next, we can define a special family of linear maps, $(f^i)_{1 \leq i \leq n}$, as follows:

$$f^i(e_j) := \delta_j^i, \quad (3.4)$$

where δ_j^i is the Kronecker delta, which is 1 if $i = j$ and 0 otherwise. Since a linear map is determined solely by its values on the basis, this condition is sufficient to define $(f^i)_{1 \leq i \leq n}$ as maps. Since we placed the index for the basis $(e_i)_{1 \leq i \leq n}$ of V below, we placed the index for $(f^i)_{1 \leq i \leq n}$ above so that Einstein's summation convention could be used later.

For any linear map $T : V \rightarrow \mathbb{R}$, if we let $T(e_i) = a_i$, then we can express $T = a_i f^i$. This can be seen by applying both sides to the basis vectors e_j . Thus, we have shown that $(f^i)_{1 \leq i \leq n}$ spans the vector space V^* . Furthermore, $(f^i)_{1 \leq i \leq n}$ is linearly independent. This is because if we assume $c_i f^i = 0$, feeding each e_j to the l.h.s. of this equation shows that all $c_j = 0$. This establishes that V^* is an n -dimensional vector space and that $(f^i)_{1 \leq i \leq n}$ is its basis. This family of linear maps is known as the *dual basis*.

3.3 Covariant vectors

With the dual space defined, the definition of a covariant vector is analogous to that of a contravariant vector.

Definition 3.3: Covariant vector

A **covariant vector** is an element of the dual vector space V^* .

This definition provides a clear and fundamental distinction between the two types of vectors: contravariant vectors live in the original vector space V , while covariant vectors live in a different, albeit related, space V^* . This immediately clarifies a common point of confusion in physics that one cannot add or subtract a contravariant vector and a covariant vector. One cannot simply add an element of V to an element of V^* , just as one cannot add a vector to a matrix. This is not an arbitrary physical dictum, but a direct consequence of the fact that they are elements of two distinct mathematical spaces.

4 The meaning of “contravariant” and “covariant”

We will now derive the component transformation rules that give these vectors their names. These rules are not axioms but are necessary consequences of our definitions when we consider a change of basis in the vector space V . Let’s consider a transformation from our original basis (e_i) to a new basis (\bar{e}_i) governed by a regular (invertible) matrix $A = (A_i^j)$

$$e_i \rightarrow \bar{e}_i = A_i^j e_j. \quad (4.1)$$

In physics, this corresponds to a coordinate transformation. That the transformed (\bar{e}_i) is actually a basis of V is a simple linear algebra problem, so the readers are strongly recommended to show it on their own.

4.1 Transformation of contravariant vector components

Let v be an arbitrary contravariant vector in V . Since it is a geometric object, the vector v itself is invariant under a change of basis. However, its components will change depending on the basis chosen. We can express v in both the old and new bases:

$$v = v^i e_i = \bar{v}^j \bar{e}_j \quad (4.2)$$

Since the index for the basis $(e_i)_{1 \leq i \leq n}$ of V was placed below, the indices for the components v^i and \bar{v}^j were written above for clarity with Einstein’s summation convention. Hereafter, the component indices of contravariant vectors will be superscripts. Now, substitute the basis transformation into this equation

$$v^i e_i = \bar{v}^j (A_j^k e_k) = (\bar{v}^j A_j^k) e_k. \quad (4.3)$$

By relabeling the dummy index i as k on the left side, we get

$$v^k e_k = (\bar{v}^j A_j^k) e_k. \quad (4.4)$$

Since the basis vectors $(e_k)_{1 \leq k \leq n}$ are linearly independent, the coefficients on both sides must be equal

$$v^k = \bar{v}^j A_j^k. \quad (4.5)$$

To solve for the new components \bar{v}^i , we multiply by the inverse matrix A^{-1}

$$(A^{-1})^i_k v^k = (A^{-1})^i_k (\bar{v}^j A_j^k) = \bar{v}^j \delta_j^i = \bar{v}^i. \quad (4.6)$$

Thus, the transformation rule for the components of a contravariant vector is

$$\bar{v}^i = (A^{-1})^i_j v^j. \quad (4.7)$$

The components of the vector transform using the *inverse* of the matrix A that was used to transform the basis vectors. This “opposite” transformation behavior is the origin of the name *contra-variant*. This relationship is necessary to ensure the physical vector v remains invariant. If the basis vectors are stretched by A , the components must shrink by A^{-1} to keep the object the same.

4.2 Transformation of covariant vector components

Now, let’s derive the transformation rule for the components of a covariant vector $T \in V^*$. First, we must determine how the dual basis $(f^i)_{1 \leq i \leq n}$ transforms. Let the new dual basis be $(\bar{f}^i)_{1 \leq i \leq n}$, defined by

$$\bar{f}^i(\bar{e}_j) = \delta_j^i. \quad (4.8)$$

We can express the old dual basis vectors in terms of the new ones, $f^k = C_l^k \bar{f}^l$. Applying this to the old basis vectors,

$$f^k(e_j) = \delta_j^k = (C_l^k \bar{f}^l)(e_j) = C_l^k \bar{f}^l((A^{-1})^m_j \bar{e}_m) = C_l^k (A^{-1})^m_j \delta_m^l = C_l^k (A^{-1})^l_j. \quad (4.9)$$

This implies that $C = A$, so

$$f^k = A_l^k \bar{f}^l. \quad (4.10)$$

The transformation for the dual basis is therefore $\bar{f}^l = (A^{-1})^l_k f^k$. (There are many indices, but do your best!!) Now, consider an arbitrary covariant vector $T \in V^*$. Like the vector v , the linear map T is an invariant object, so its representation can be written in both bases

$$T = a_i f^i = \bar{a}_j \bar{f}^j. \quad (4.11)$$

The indices for the components a_i are placed below to comply with the summation convention. Substitute the transformation for the dual basis

$$a_i f^i = \bar{a}_j ((A^{-1})^j_k f^k) = (\bar{a}_j (A^{-1})^j_k) f^k. \quad (4.12)$$

By linear independence of the dual basis vectors $(f^k)_{1 \leq k \leq n}$, we equate the coefficients

$$a_k = \bar{a}_j (A^{-1})^j_k. \quad (4.13)$$

Multiplying by the matrix A to solve for \bar{a}_i

$$a_k A_i^k = \bar{a}_j (A^{-1})^j_k A_i^k = \bar{a}_j \delta_i^j = \bar{a}_i. \quad (4.14)$$

Table 1. Comparison of contravariant and covariant vectors

Feature	Contravariant Vector	Covariant Vector
Belongs to Space	Vector Space V	Dual Vector Space V^*
Nature of Object	Geometric “arrow”	Linear map from V to \mathbb{R}
Basis	$(e_i)_{1 \leq i \leq n}$	Dual Basis $(f^i)_{1 \leq i \leq n}$
Component Index	Upper (e.g., v^i)	Lower (e.g., a_i)
Basis Transformation	$\bar{e}_i = A_i^j e_j$	$\bar{f}^i = (A^{-1})_j^i f^j$
Component Transformation	$\bar{v}^i = (A^{-1})_j^i v^j$	$\bar{a}_i = a_j A_i^j$

The transformation rule for the components of a covariant vector is

$$\bar{a}_i = a_j A_i^j. \quad (4.15)$$

The components of the covariant vector transform using the *same* matrix A as the basis vectors e_i . This shared transformation behavior is the origin of the name *co-variant*.

The following table summarizes these fundamental properties. In other words, the physics-style definitions of contravariant and covariant vectors are those that define how the *components* ($\in \mathbb{R}$) of the elements of V and the dual space V^* transform when the basis of the vector space V is transformed. The indices for the components of contravariant vectors are upper, and for covariant vectors are lower; this is a unified convention that is the same in any physics book you read.

5 Tensors

The concepts of contravariant and covariant vectors can be naturally generalized to tensors.

5.1 Definition of a tensor

Finally, we define the **tensor product space**, which is the set of tensors. What is called a “tensor” in physics is an element of a special case of this tensor product space.

Definition 5.1: Tensor product

The **tensor product** of vector spaces V_1 and V_2 is a pair $(V_1 \otimes V_2, \phi)$ consisting of a vector space $V_1 \otimes V_2$ and a map $\phi : V_1 \times V_2 \rightarrow V_1 \otimes V_2$ that satisfies the following *universality* (*):

(*) For any vector space W and any bilinear map $f : V_1 \times V_2 \rightarrow W$, there exists a unique linear map $g : V_1 \otimes V_2 \rightarrow W$ such that $f = g \circ \phi$:

$$\begin{array}{ccc} V_1 \times V_2 & \xrightarrow{\forall f} & W \\ \phi \downarrow & \nearrow \exists! g & \\ V_1 \otimes V_2 & & \end{array} \quad (5.1)$$

This definition is the most mathematically general definition of a tensor product space, but to be honest, its meaning seems completely lost on those from outside a pure mathematics background. So let's try to rephrase it in a more digestible way.

“Definition”: Tensor product (intuitive)

Let V_1 be a k -dimensional vector space with basis $(e_i^1)_{1 \leq i \leq k}$ and V_2 be an l -dimensional vector space with basis $(e_j^2)_{1 \leq j \leq l}$. Their tensor product, $V_1 \otimes V_2$, is a kl -dimensional vector space whose basis is given by the set of all formal products $(e_i^1 \otimes e_j^2)_{1 \leq i \leq k, 1 \leq j \leq l}$.

This definition lacks some mathematical rigor, but it is intuitive, so it is sufficient for handling tensors in physics. For details on how this is derived from the more abstract definition, please refer to an advanced linear algebra textbook or a book on commutative ring theory [1, 2]. The tensor product can also be defined for three or more vector spaces. This construction allows us to define the tensors used in physics.

Definition 5.2:

A **tensor of type (k, l)** , also known as a **k -th contravariant, l -th covariant tensor**, is an element of the tensor product space

$$\underbrace{V \otimes \dots \otimes V}_{k \text{ times}} \otimes \underbrace{V^* \otimes \dots \otimes V^*}_{l \text{ times}}. \quad (5.2)$$

From this definition, it is clear that a contravariant vector is simply a tensor of type $(1, 0)$, and a covariant vector is a tensor of type $(0, 1)$. This provides a unified framework for all these objects.

5.2 Tensor components and their transformation

The basis for the space of (k, l) -tensors is formed by the tensor products of the basis vectors of V and V^*

$$(e_{i_1} \otimes \dots \otimes e_{i_k} \otimes f^{j_1} \otimes \dots \otimes f^{j_l})_{1 \leq i_1, \dots, i_k, j_1, \dots, j_l \leq n} \quad (5.3)$$

For example, in the case where $\dim V = n = 2$, the basis for the space of $(1, 1)$ -tensors, $V \otimes V^*$, consists of the four vectors $(e_1 \otimes f^1, e_1 \otimes f^2, e_2 \otimes f^1, e_2 \otimes f^2)$. An arbitrary (k, l) -tensor T can be expanded in this basis

$$T = T_{j_1 \dots j_l}^{i_1 \dots i_k} e_{i_1} \otimes \dots \otimes e_{i_k} \otimes f^{j_1} \otimes \dots \otimes f^{j_l}. \quad (5.4)$$

The coefficients $T_{j_1 \dots j_l}^{i_1 \dots i_k} \in \mathbb{R}$ are the components of the tensor. In physics, it is common to refer to this collection of components $(T_{j_1 \dots j_l}^{i_1 \dots i_k})_{1 \leq i_1, \dots, i_k, j_1, \dots, j_l \leq n}$ as the tensor itself. The transformation rule for these components can be derived directly from the invariance of the tensor object T and the transformation rules for the basis vectors e_i and dual basis vectors f^j . The result follows from eqs. (4.1) & (4.10) (the indices are tough!):

$$\bar{T}_{j_1 \dots j_l}^{i_1 \dots i_k} = (A^{-1})_{m_1}^{i_1} \dots (A^{-1})_{m_k}^{i_k} A_{j_1}^{n_1} \dots A_{j_l}^{n_l} T_{n_1 \dots n_l}^{m_1 \dots m_k} \quad (5.5)$$

Each contravariant (upper) index transforms with a factor of (A^{-1}) , and each covariant (lower) index transforms with a factor of A . This is precisely the transformation rule that physics textbooks often use as the *definition* of a tensor. Here, we see it is a direct consequence of the more fundamental definition of a tensor as an element of a tensor product space.

6 Tensors as multilinear maps

There is another powerful and equivalent way to view tensors: as multilinear maps. This perspective often clarifies their role in physical theories.

A covariant vector $T \in V^*$ is, by definition, a linear map from V to \mathbb{R} . We can generalize this:

- A $(0, l)$ -*tensor* (a purely covariant tensor) can be viewed as a multilinear map that takes l vectors from V as input and produces a scalar in \mathbb{R} ,

$$T : \underbrace{V \times \cdots \times V}_{l \text{ times}} \rightarrow \mathbb{R}. \quad (6.1)$$

- A $(k, 0)$ -*tensor* (a purely contravariant tensor) can be viewed as a linear map $V^* \rightarrow \mathbb{R}$. Specifically, for $v \in V$ and $T \in V^*$, we define $v(T) := T(v)$. (To define this with mathematical rigor, the concept of a *double dual space* is used, which establishes that $V^{**} \cong V$). This can be extended to a multilinear map that takes k covectors from V^* as input and produces a scalar,

$$T : \underbrace{V^* \times \cdots \times V^*}_{k \text{ times}} \rightarrow \mathbb{R}. \quad (6.2)$$

- A general (k, l) -*tensor* can be viewed as a multilinear map that takes k covectors and l vectors as input and produces a scalar,

$$T : \underbrace{V^* \times \cdots \times V^*}_{k \text{ times}} \times \underbrace{V \times \cdots \times V}_{l \text{ times}} \rightarrow \mathbb{R}. \quad (6.3)$$

This functional perspective turns a tensor from a static object with many components into a “machine” that processes vectors and covectors to yield a scalar. This is precisely how tensors are used to construct coordinate-invariant quantities (scalars) in physics, such as the Lagrangian or the spacetime interval.

7 Application in physics: the Minkowski metric

A crucial example of a tensor in physics is the Minkowski metric of special relativity.

7.1 The Minkowski metric as a $(0, 2)$ -tensor

In special relativity, spacetime is modeled as a 4-dimensional vector space $V = \mathbb{R}^{1,3}$, known as Minkowski space. The Minkowski metric η is a $(0, 2)$ -tensor. Using the basis $(e_\mu)_{0 \leq \mu \leq 3}$ and dual basis $(f^\mu)_{0 \leq \mu \leq 3}$, the metric tensor can be written as

$$\eta = \eta_{\mu\nu} f^\mu \otimes f^\nu. \quad (7.1)$$

The components $\eta_{\mu\nu}$ form a diagonal matrix, typically with the signature $(+1, -1, -1, -1)$ or $(-1, +1, +1, +1)$.¹ Let's use the latter

$$(\eta_{\mu\nu}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (7.2)$$

In other words,

$$\eta = -f^0 \otimes f^0 + f^1 \otimes f^1 + f^2 \otimes f^2 + f^3 \otimes f^3. \quad (7.3)$$

Viewing η as a multilinear map, it takes two vectors $v = v^\mu e_\mu$ and $w = w^\nu e_\nu$ from V and produces their inner product

$$\eta(v, w) = (\eta_{\alpha\beta} f^\alpha \otimes f^\beta)(v^\mu e_\mu, w^\nu e_\nu) = \eta_{\alpha\beta} f^\alpha(v^\mu e_\mu) f^\beta(w^\nu e_\nu) = \eta_{\alpha\beta} v^\mu \delta_\mu^\alpha w^\nu \delta_\nu^\beta = \eta_{\mu\nu} v^\mu w^\nu. \quad (7.4)$$

This reproduces the familiar formula for the Minkowski inner product, $-v^0 w^0 + v^1 w^1 + v^2 w^2 + v^3 w^3$.

7.2 Raising and lowering indices

In physics literature, one frequently encounters the practice of “raising” and “lowering” indices using the metric tensor, for example, writing $v_\mu = \eta_{\mu\nu} v^\nu$. Our framework provides a rigorous mathematical meaning for this operation. It is not just a notational trick but a well-defined map between the vector space V and its dual V^* , facilitated by the metric. The metric tensor η , being a $(0, 2)$ -tensor, is a machine that takes two vectors as input. If we feed it only one vector, say $v \in V$, what remains is an object that is waiting for a second vector to produce a scalar. This object, denoted $\eta(\cdot, v)$, is a linear map from V to \mathbb{R} , which means it is a covariant vector—an element of V^* .

$$\eta(\cdot, v) = (\eta_{\mu\nu} v^\nu) f^\mu \in V^* \quad (7.5)$$

Let's find the components of this new covariant vector. Let's call them $v_\mu := \eta_{\mu\nu} v^\nu$. By definition, v_μ are the components of the covector $\eta(\cdot, v)$ in the dual basis $(f_{0 \leq \mu \leq 3}^\mu)$. The action of this covector on a basis vector e_ν is

$$(\eta(\cdot, v))(e_\nu) = \eta(e_\nu, v) = \eta_{\alpha\beta} e_\nu^\alpha v^\beta = \eta_{\alpha\beta} \delta_\nu^\alpha v^\beta = \eta_{\nu\beta} v^\beta. \quad (7.6)$$

¹General relativity researchers prefer the former, while the high energy physicists would pick the latter.

On the other hand, the components v_μ are defined by the expansion $\eta(\cdot, v) = v_\mu f^\mu$. Its action on e_ν is

$$(v_\mu f^\mu)(e_\nu) = v_\mu f^\mu(e_\nu) = v_\mu \delta_\nu^\mu = v_\nu. \quad (7.7)$$

Equating the two results, we find the components of the new covector:

$$v_\nu = \eta_{\nu\mu} v^\mu \quad (7.8)$$

This provides a rigorous justification for the index-lowering operation. To summarize, the Minkowski metric η is a $(0, 2)$ -tensor, so it is originally a multilinear map $\eta : V \times V \rightarrow \mathbb{R}$, but by taking only one argument, it can also be regarded as a map $\eta : V \rightarrow V^*$. “Raising and lowering indices” was a map that sends elements of V to elements of V^* . In Euclidean space, where the metric components form the identity matrix, the components of a vector and its dual are identical, so the distinction is often ignored. In Minkowski space and general relativity, the metric is non-trivial, making the distinction between contravariant and covariant objects essential.

8 Conclusion

This article has sought to build an understanding of vectors and tensors from a mathematical foundation. The core concepts can be summarized as follows:

- A *contravariant vector* is an element of a vector space V .
- A *covariant vector* is an element of the dual vector space V^* .
- A *tensor* is an element of a tensor product space constructed from V and V^* .

The famous transformation properties that characterize these objects in physics are not fundamental definitions but are merely consequences of these more basic algebraic structures. This perspective reveals that the rules of tensor calculus are not arbitrary but are the necessary mathematical language for expressing physical laws in a way that is independent of any specific choice of coordinates. Physicists are often brilliant at generating novel ideas and physical intuition, while mathematicians excel at abstracting and systematizing diverse concepts into a coherent, logical framework. By adopting a mathematical viewpoint, we can often bring greater clarity and structure to our physical knowledge, transforming rote rules into logical necessities.

References

- [1] A. Yukie, Algebra 2: Rings, Fields, and Galois Theory. Nippon Hyoron Sha, 2010.
- [2] M. F. Atiyah and I. G. Macdonald, Introduction to commutative algebra. CRC Press, 2018.