Covariance and contravariance

tfb / washdarb

20150312
(in memory of Terry Pratchett)

Contents

1 The vector space \( T \) and its bases .................. 2
2 The dual space \( T^* \) and its bases .................. 4
3 Summary for vectors and one-forms ............... 7
4 Tensors .................................................. 8
5 Summary for tensors .................................. 10
6 Remembering the rules: row and column matrices .. 11
7 Notation, contravariance and covariance ........... 14
8 Things to remember about square matrices ........ 16
9 Bibliography ............................................. 17

This is a small attempt to describe vectors and one-forms (traditionally known to physicists as ‘contravariant and covariant vectors’) in a slightly more modern way than the awful approach often still taught, while not dragging in any significant differential geometry. An
introduction to differential geometry suitable for physicists can be
found in [1] and other books cited there, to which this document is
heavily indebted.

The aim specifically is to show how the transformation rules
for components arise, rather than simply learning them by rote as
physicists usually do. The transformation rules are summarised in
section 3 (vectors and one-forms) and section 5 (tensors).

1 The vector space $T$ and its bases

Let $T$ be an $n$-dimensional vector space defined over $\mathbb{R}$. Write
elements of this space as $v \in T$. $n$ should be finite, but all of this
will work over fields other than $\mathbb{R}$: I'm just sticking to $\mathbb{R}$ because it's
the interesting field in GR.

Pick a basis $\{e_i, i = 1, \ldots, n\}$ for $T$. Then we can express any
vector $v \in T$ as

$$v = \sum_{i=1}^{n} v^i e_i$$

(Here, and generally, superscripts just mean indices, not powers.)

The Einstein summation convention. We will have a lot of sums of
the form $\sum_{i=1}^{n} a^i b_i$, so we adopt a convention that repeated indices,
one raised and one lowered, are summed over. In other words

$$v^i e_i \equiv \sum_{i=1}^{n} v^i e_i$$

We'll use this convention from here on, except where specified.
Now pick a new basis for $T$, $\{e'_i\}$. We can express this basis in terms of the old basis as

$$e'_i = \Lambda^i_j e_j \quad (1)$$

where the matrix $\Lambda$, with elements $\Lambda^i_j$, is non-singular. Note that the $\{\Lambda^i_j\}$ are not the components of a tensor. The inverse relationship is

$$e_i = (\Lambda^{-1})^i_j e'_j \quad (2)$$

Now any $v \in T$ can be expressed both as $v = v^i e_i$ and as $v = v^i e'_i$, where the $\{v^i\}$ are the components in the $\{e'_i\}$ basis.

$$v = v^i e_i = v^i (\Lambda^{-1})^i_j e'_j = v^i e'_i$$

Hence

$$v^i = v^i (\Lambda^{-1})^i_j \quad (3)$$

This equation is arranged to make it look like right-multiplication by $\Lambda^{-1}$: see section 6 for why this is natural. If you want to think of it as left-multiplication you need a transpose: $v^i = ((\Lambda^T)^{-1})^i_j v^j$.

Similarly

$$v = v^i e'_i = v^i \Lambda^i_j e_j = v^i e'_i$$
Hence

(4) \[ v^i = v'^i \Lambda_j^j \]
(or \( v^i = (\Lambda^T)^i_j v'^j \)).

It’s worthwhile checking (3) and (4):

\[

v^i = v'^i \Lambda_j^j \quad \text{from (4)} \\
= v^k (\Lambda^{-1})_k^j \Lambda_j^i \quad \text{from (3)} \\
= v^k (\Lambda^{-1} \Lambda)_k^i \\
= v^k \delta_k^i \\
= v^i

\]

So this is all fine. Finally note that the occurrence of \( \Lambda^{-1} \) in the transformation rule is why vector components were called ‘contravariant’.

2 The dual space \( T^* \) and its bases

Consider linear functions \( T \to \mathbb{R} \). We’ll write such functions as \( \omega \), and use the notation \( \langle \omega, \mathbf{v} \rangle \) for function application\(^1\). Linearity means that \( \langle \omega, \alpha \mathbf{v} + \beta \mathbf{u} \rangle = \alpha \langle \omega, \mathbf{v} \rangle + \beta \langle \omega, \mathbf{u} \rangle \), \( \alpha, \beta \in \mathbb{R} \). We can then define addition of these linear functions and multiplication by

\(^1\)We could also write function application more conventionally as \( \omega(\mathbf{v}) \), and this is the notation used in [1], but this Dirac-inspired notation seems more appropriate: see section 7 below.
Scalars: \( \langle \alpha \omega + \beta \varphi, \nu \rangle = \alpha \langle \omega, \nu \rangle + \beta \langle \varphi, \nu \rangle \). Define the basis linear functions \( \{ \mathbb{e}^i : T \to \mathbb{R}, i = 1, \ldots, n \} \) by \( \langle \mathbb{e}^i, e_j \rangle = \delta^i_j \). Then any linear function \( \omega \) can be written

\[
\omega = \sum_{i=1}^{n} \omega_i \mathbb{e}^i
\]

and

\[
\langle \omega, \nu \rangle = \omega_i \langle \mathbb{e}^i, \nu \rangle = \omega_i \langle \mathbb{e}^i, \nu e_j \rangle = \omega_i \nu_j \langle \mathbb{e}^i, e_j \rangle = \omega_i \nu_j \delta^i_j = \omega_i \nu^i.
\]

Clearly these functions form a vector space over \( \mathbb{R} \): the dual space to \( T \), called \( T^\ast \). \( \{ \mathbb{e}^i \} \) is the dual basis induced in \( T^\ast \) by the basis \( \{ e_i \} \) in \( T \). Since this space consists of linear functions, or forms, of one argument, we will call \( \omega \in T^\ast \) a one-form.

We want to know what happens to the components of a one-form \( \omega \) when we change the basis of \( T \) from \( \{ e_i \} \) to \( \{ e'_i \} \) and make the corresponding basis change in \( T^\ast \). Well, first let’s look at the change of basis in \( T^\ast \). Corresponding to the basis \( \{ e'_i \} \) for \( T \) there will be a basis \( \{ e'^i \} \) for \( T^\ast \), and \( \langle e'^i, e'_j \rangle = \delta^i_j \). Let

\[
\mathbb{e}^i = \mathbb{e}'^i M^i_j \quad \text{(or } \mathbb{e}^i = (M^i_j) \mathbb{e}'^i \text{)}
\]

5
\(M\) is a nonsingular matrix since both \(\{\mathbf{e}^i\}\) and \(\{\mathbf{e}^i\}'\) are bases. Now, from above and (1),

\[
\langle \mathbf{e}^i, \mathbf{e}^j' \rangle = \langle \mathbf{e}^k M^i_k, \Lambda^i_j \mathbf{e}_i \rangle \\
= \Lambda^i_j M^i_k \langle \mathbf{e}^k, \mathbf{e}_j \rangle \\
= \Lambda^i_j M^i_k \delta^k_l \\
= \Lambda^i_j M^i_j
\]

but

\[
\langle \mathbf{e}^i', \mathbf{e}^j' \rangle = \delta^i_j
\]

so

\[
\Lambda^i_j M^i_k = \delta^i_j
\]

or

\[
M = (\Lambda)^{-1}
\]

Hence

(5) \(\mathbf{e}^i'' = \mathbf{e}^i (\Lambda^{-1})^i_j\)

and

(6) \(\mathbf{e}^i' = \mathbf{e}^j' \Lambda^i_j\)

Both of these could also be written in terms of \(\Lambda^T\) to make it look like left-multiplication instead of right-multiplication: see section 6 again.
Now
\[
\omega = \omega_i e^i = \omega_i e'^i / \Lambda^i_j = \omega_j e_j'
\]
So
\[
(7) \quad \omega'_i = \Lambda^i_j \omega_j
\]

Similarly
\[
\omega = \omega'_i e'^i = \omega'_i (\Lambda^{-1})^i_j = \omega_j e_j'
\]
So
\[
(8) \quad \omega_i = (\Lambda^{-1})^i_j \omega'_j
\]

3 Summary for vectors and one-forms

In summary the following equations give rules for what happens corresponding to a change of basis in \( T \) from \( \{ e_i \} \) to \( \{ e'_i \} \), given by a matrix \( \Lambda \):

- **basis vectors:** \( e'_i = \Lambda^i_j e_j, \quad e_i = (\Lambda^{-1})^i_j e_j \) – \(1), (2)\;
- **vector components:** \( v'^i = v^i (\Lambda^{-1})^i_j, \quad v^i = v^i_j \Lambda^j_i \) – \(3), (4)\;
- **basis one-forms:** \( e'^i = e^i (\Lambda^{-1})^i_j, \quad e^i = e^i_j \Lambda^j_i \) – \(5), (6)\;
one-form components: $\omega'_i = \Lambda^j_i \omega_j$, $\omega_i = (\Lambda^{-1})^j_i \omega'_j - (7), (8)$.

It is important to understand three things.

**Vectors and one-forms are not simply their components:** they are geometrical objects living in the vector space $T$ or the dual space $T^*$.

**Basis vectors and one-forms change:** the two sets of basis vectors in each space which we have considered are different sets of vectors.

**Other vectors and one-forms do not change:** all that is changing is the coefficients of the changed basis vectors or one-forms used to express the vector or one-form: it is the same object.

### 4 Tensors

We’ve seen that one-forms are linear functions from $T \to \mathbb{R}$, and equivalently vectors are linear functions from $T^* \to \mathbb{R}$: tensors are simply a generalisation of this. A tensor is a multilinear function from vectors and one-forms to $\mathbb{R}$. In particular, if $U$ is an $\binom{N}{M}$ tensor$^2$ this means it is a multilinear function

$$U : \underbrace{T^* \times \cdots \times T^*}_{N \text{ copies}} \times \underbrace{T \times \cdots \times T}_{M \text{ copies}} \to \mathbb{R}$$

That’s all a tensor is. Two important types of tensor are one-forms, which are $\binom{0}{1}$ tensors, and vectors, which are $\binom{1}{0}$ tensors.

$^2$I’d like to call this $T$ but obviously I can’t without doing some typeface trick.
We can write a tensor as a function, \( U(\cdot, \ldots; \cdot, \ldots) \), where the first set of arguments are one-forms and the second are vectors. Linearity then means that \( U(\alpha \omega + \beta \sigma, \ldots; \cdot, \ldots) = \alpha U(\omega, \ldots; \cdot, \ldots) + \beta U(\sigma, \ldots; \cdot, \ldots) \), and the same for all the other arguments.

A little bit of thought\(^3\) should convince you that a function like this can be written as a linear combination:

\[
U(\cdot, \ldots; \cdot, \ldots) = U_{i_1 \cdots i_N}^{j_1 \cdots j_M} \langle e_{i_1}, \cdot \rangle \cdots \langle e_{i_N}, \cdot \rangle \langle e^{j_1}, \cdot \rangle \cdots \langle e^{j_M}, \cdot \rangle
\]

So the question is: how do the \( \{ U_{i_1 \cdots i_N}^{j_1 \cdots j_M} \} \) transform under a change of basis? This is almost obvious in fact, but we'll work it out for a \( (1,1) \) tensor to make it clear. First of all let's write everything in the unprimed bases:

\[
U(\omega, v) = U_{i}^{j} \langle e_{i}, \omega \rangle \langle e^{j}, v \rangle
\]

\[
= U_{i}^{j} \langle e_{i}, \omega_k e^k \rangle \langle e^{j}, v^l e_l \rangle
\]

\[
= U_{i}^{j} \omega_k v^l \langle e_{i}, e^k \rangle \langle e^{j}, e_l \rangle
\]

\[
= U_{i}^{j} \omega_k v^l \delta_i^k \delta^j_l
\]

\[
= U_{i}^{j} \omega_i v^j
\]

which is as we would expect\(^4\). In the primed basis it immediately follows that \( U(\omega, v) = U'_{i}^{j} \omega'_{i} v'^{j} \) and hence

\[
U'_{i}^{j} \omega'_{i} v'^{j} = U_{i}^{j} \omega_{i} v^{j}
\]

\(^3\)Every time I write this I have to convince myself over again that it is obvious.

\(^4\)This is the ‘little bit of thought’ you need to convince yourself, in fact.
But from (4) and (8), \( \omega_l = (\Lambda^{-1})^i_l \omega'_l \) and \( \nu^m = \nu^j_l \Lambda^m_j \), so

\[
U^l_m \omega_l \nu^m = U^l_m (\Lambda^{-1})^i_l \Lambda^m_j \omega'_l \nu^j_l
\]

and hence

\[
(9) \quad U'^i_j = \Lambda^m_j U^l_m (\Lambda^{-1})^i_l
\]

And similarly

\[
(10) \quad U^i_j = (\Lambda^{-1})^m_j U^l_m \Lambda^i_l
\]

It is clear that this generalises to arbitrary tensor ranks.

5 Summary for tensors

(9) and (10) give the transformation rules for tensor components, and they are what you would expect: raised indices are like vector components:

\[
U'^{...i...} = U^{...i...} (\Lambda^{-1})^i... \quad \text{like (3)}
\]

\[
U^{...i...} = U'^{...j...} \Lambda^i_j... \quad \text{like (4)}
\]

and lowered indices are like one-form components:

\[
U'^{...i...} = \Lambda^i_j... U^{...j...} \quad \text{like (7)}
\]

\[
U^{...i...} = (\Lambda^{-1})^i_j... U'^{...j...} \quad \text{like (8)}
\]
6 Remembering the rules: row and column matrices

[The start of this section does not use the Einstein summation convention, as it’s trying to be conventional matrix algebra.] It’s obvious that the space of $n \times 1$ matrices – column matrices with $n$ rows – forms an $n$-dimensional vector space. Well, one-forms are linear functions from vectors to scalars, so for a column matrix such as

$$
\begin{pmatrix}
c_1 \\
n \\
c_n
\end{pmatrix}
$$

this is an expression such as $\sum_i c_i r_i$ for some scalars $\{r_i\}$. In other words, one-forms are row matrices:

$$
(r_1 \cdots r_n) \times \begin{pmatrix}
c_1 \\
n \\
c_n
\end{pmatrix} = \sum_{i=1}^n r_i c_i
$$

Equivalently, column matrices serve as linear functions from row matrices to the reals.

So, what about square matrices? Well

$$
\begin{pmatrix}
A_{11} & \cdots & A_{1n} \\
\vdots & \ddots & \vdots \\
A_{n1} & \cdots & A_{nn}
\end{pmatrix} \times \begin{pmatrix}
c_1 \\
n \\
c_n
\end{pmatrix} = \begin{pmatrix}
\sum_i A_{1i} c_i \\
\vdots \\
\sum_i A_{ni} c_i
\end{pmatrix}
$$

So a matrix is a $\left( \begin{smallmatrix} 1 \\ 1 \end{smallmatrix} \right)$ tensor, or a linear map from column matrices to column matrices. We can also right-multiply one-forms – row
matrices – by square matrices:

\[(r_1 \cdots r_n) \times \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} = (\sum_i r_i A_{i1} \cdots \sum_i r_i A_{in})\]

And again we can see that matrices are \(\binom{1}{1}\) tensors.

So, now we can do a slightly devious thing. Using the Einstein summation convention again, remember that \(v^i = v^i e_i\). Well, we can write this like this:

\[(11) \quad v = (v^1 \cdots v^n) \times \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}\]

The ‘slightly devious’ bit is that the ‘scalars’ in (11) are either vectors in \(T\) or reals.

And now consider a nonsingular matrix \(\Lambda\) whose elements are \(\{\Lambda^i_j\}\), and its inverse \(\Lambda^{-1}\) with elements \(\{(\Lambda^{-1})^i_j\}\). We can obviously put \(I = \Lambda^{-1}\Lambda\) into the middle of the expression for \(v\) in (11):

\[
\begin{align*}
\quad v &= (v^1 \cdots v^n) \times \begin{pmatrix} (\Lambda^{-1})^1_1 & \cdots & (\Lambda^{-1})^1_n \\ \vdots & \ddots & \vdots \\ (\Lambda^{-1})^n_1 & \cdots & (\Lambda^{-1})^n_n \end{pmatrix} \\
& \quad \times \begin{pmatrix} \Lambda^1_1 & \cdots & \Lambda^1_n \\ \vdots & \ddots & \vdots \\ \Lambda^n_1 & \cdots & \Lambda^n_n \end{pmatrix} \times \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}
\end{align*}
\]
Well, matrix multiplication is associative, so we can do the multiplications within the two lines above first, to get

\[ v = (v^1(\Lambda^{-1})_1 \cdots v^\mu(\Lambda^{-1})_\mu) \times \begin{pmatrix} \Lambda^1_1 \omega_1 \\ \vdots \\ \Lambda^\mu_\mu \omega_\mu \end{pmatrix} \]

And the transformation rules for vector components can be read from this directly.

We can then do the same thing for one-forms, by writing \( \omega = \omega_i e^i \) as \( \omega = \omega_i e^i \):

\[ (12) \quad \omega = (e^1 \cdots e^\mu) \times \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_\mu \end{pmatrix} \]

which leads to the transformation rules for one-form components.

There is nothing new in (11) and (12) that is not in section 1 and section 2: the purpose of them is that they give you an easy way of remembering the transformation rules in terms of matrix algebra, without getting lost down the usual path of thinking of everything as ‘matrices with special transformation rules’.

- Basis vectors are like column matrices, so left-multiply by \( \Lambda \), or \( e'_i = \Lambda^j_i e_j \).
- Vector components are like row matrices, so right-multiply by \( \Lambda^{-1} \), or \( v'^i = v^\mu(\Lambda^{-1})^\mu_i \).
- Basis one-forms are like row matrices, so right-multiply by \( \Lambda^{-1} \), or \( e''_i = e^i(\Lambda^{-1})^i_j \).
one-form components are like column matrices, so left-multiply by \( \Lambda \), or \( \omega_i = \Lambda^i_j \omega_j \).

Finally, you can remember which are row and which are column matrices by where the indices are: if the index is raised it’s a row, if lowered it’s a column.

### 7 Notation, contravariance and covariance

I have used \( T \) as the name of my vector space, because, in real life it is the tangent space to some manifold (this has meant I’ve had to call my sample tensor \( U \)). The notation for vectors and one-forms varies a lot, as do index conventions and so on: I’ve used bars and tildes under letters because it is easy to hand write. I’m not sure why I also used Greek letters for one-forms (but not basis one-forms): Schutz uses Greek for both but I find it much easier if all basis vectors and one-forms are called \( e \). One-forms are also called ‘covectors’.

I don’t think there’s a good answer to where primes go when changing basis: \( e'_i \) is a different vector to \( e_i \), but the \( \{v^i\} \) are the components of the same vector, \( \tilde{v} \), as the \( \{v^i\} \), just on a different basis. Similarly \( \{U^{i\cdots}\} \) are the components of the same tensor as \( \{U^{i\cdots}\} \) are. Schutz primes the indices – \( e_i \) – which I find even more confusing, which is why I have not done it.

It should be clear that the two spaces are equivalent: this is why I have used the \( \langle \cdot, \cdot \rangle \) notation: the more conventional notation \( \tilde{\omega}(\tilde{v}) \) obscures the symmetry. Indeed we could also write \( \langle \tilde{v}, \omega \rangle \). This notation is intentionally reminiscent of Dirac’s bra-ket notation: in quantum mechanics kets are vectors and bras are one-forms, although
things are more complicated since the space is not finite-dimensional in general.

It is important to understand that there is no bijection between a one-form and a corresponding vector here: \( T \) and \( T^* \) are real vector spaces with the same dimension, and a basis in one induces a basis in the other, but we cannot take a one-form \( \omega \) and map it to a corresponding vector \( v \). In component terms there is no index raising or lowering operation. To do that needs a metric, which we do not have.

Vector components were called ‘contravariant’ because their transformation rule involves \( \Lambda^{-1} \); one-form components were called ‘covariant’ because their rule involves \( \Lambda \).

The traditional approach when teaching this material to physicists seems, still, to involve getting completely obsessed with the transformation rules and essentially defining vectors (‘contravariant vectors’) and one-forms (‘covariant vectors’) as a bundle of numbers with some special magic transformation rules: I hope I’ve convinced at least some people that that approach serves mainly to obfuscate what is going on: vectors and one-forms are geometrical objects, components are merely how we talk about them, and in fact merely one way of doing even that.
8 Things to remember about square matrices

Some useful things, a few of which are needed above.

\[
A = \begin{pmatrix}
A_{11} & A_{12} & \cdots & A_{1n} \\
A_{21} & A_{22} & \cdots & A_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
A_{n1} & A_{n2} & \cdots & A_{nn}
\end{pmatrix}
\]

index order

\[
l_{ij} = \delta_{ij} = \begin{cases} 
1, & i = j \\
0, & i \neq j
\end{cases}
\]

Kronecker delta

\[
(AB)_{ij} = \sum_{k=0}^{n} A_{ik} B_{kj}
\]

matrix–matrix product

\[
(AV)_i = \sum_{j=0}^{n} A_{ij} V_j
\]

matrix–vector product

\[
(A^T)_{ij} = A_{ji}
\]

transpose

\[
(AB)^T = B^T A^T
\]

transpose of product

\[
(AB)^{-1} = B^{-1} A^{-1}
\]

inverse of product

\[
(A^T)^{-1} = (A^{-1})^T
\]

inverse and transpose

Formerly I used the last of these a fair bit in an attempt to make everything look like matrix–vector products, preferring \((\Lambda^T)^j v^j\) to \(\Lambda^i v^i\) for instance: I’m not doing that now.


9 Bibliography