

Problem Set 1: Linear Algebra

1 Dual Basis (5 points)

Let

$$\vec{e}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \vec{e}_2 = \begin{pmatrix} 4 \\ 7 \end{pmatrix} \quad (1)$$

be a basis for $\mathbb{R}^2$, where the column vectors correspond to their components in the standard orthonormal Cartesian basis, denoted by $\hat{x}_i$. Find the basis covectors $\{\vec{e}^{*1}, \vec{e}^{*2}\}$, which are dual to $\{\vec{e}_1, \vec{e}_2\}$, written in terms of the usual dual basis $\hat{x}^{*i}$, where $\hat{x}^{*i}(\hat{x}_j) = \delta_j^i$.

SOLUTION:

We know that

$$\vec{e}^{*i}(\vec{e}_j) = \delta_j^i \quad (2)$$

Write $\vec{e}_i = a^j_i \hat{x}_j$, and $\vec{e}^{*i} = f^i_j \hat{x}^{*j}$, where f^i_j are the components of $\vec{e}^{*i}$ in the standard basis. By definition,

$$\vec{e}^{*i}(\vec{e}_j) = \delta_j^i. \quad (3)$$

But we also have

$$\vec{e}^{*i}(\vec{e}_j) = f^i_l \hat{x}^{*l}(a^k_j \hat{x}_k) = f^i_l a^k_j \delta^l_k = f^i_k a^k_j. \quad (4)$$

Therefore,

$$f^i_k a^k_j = \delta_j^i \implies f^i_k = (a^{-1})^i_k. \quad (5)$$

But as a matrix,

$$a = \begin{pmatrix} 2 & 4 \\ 3 & 7 \end{pmatrix} \implies a^{-1} = \frac{1}{2} \begin{pmatrix} 7 & -4 \\ -3 & 2 \end{pmatrix} \quad (6)$$

Thus,

$$\vec{e}^{*1} = \frac{7}{2} \hat{x}^{*1} - 2 \hat{x}^{*2} \quad \vec{e}^{*2} = -\frac{3}{2} \hat{x}^{*1} + \hat{x}^{*2}. \quad (7)$$

2 (SG 10.10) Quotient Theorem (5 points)

Suppose that you have come up with some recipe for generating an array of numbers T^{ijk} in any coordinate frame, and want to know whether these numbers are the components of a triply contravariant tensor. Suppose further that you know that, given the components of a_{ij} of an arbitrary doubly covariant tensor, the numbers

$$T^{ijk} a_{jk} = v^i \quad (8)$$

transforms as the components of a contravariant vector. Show that T^{ijk} does indeed transform as a triply contravariant tensor. (The natural generalization of this result to arbitrary tensor types is known as the **quotient theorem**.)

SOLUTION:

Under a change of basis $\vec{e}'_j = \alpha^i_j \vec{e}_i$, we have

$$a_{jk} = \alpha^{j'}_j \alpha^{k'}_k a_{j'k'}, \quad v^i = (\alpha^{-1})^i_{i'} v^{i'}. \quad (9)$$

Therefore, we have

$$\begin{aligned} T^{ijk} \alpha^{j'}_j \alpha^{k'}_k a_{j'k'} &= (\alpha^{-1})^i_{m'} v^{m'} \implies \alpha^{i'}_i T^{ijk} \alpha^{j'}_j \alpha^{k'}_k a_{j'k'} = \alpha^{i'}_i (\alpha^{-1})^i_{m'} v^{m'} \\ &\implies \alpha^{i'}_i \alpha^{j'}_j \alpha^{k'}_k T^{ijk} a_{j'k'} = \delta^{i'}_{m'} v^{m'} = v^{i'} \end{aligned} \quad (10)$$

Defining $T^{i'j'k'} = \alpha^{i'}_i \alpha^{j'}_j \alpha^{k'}_k T^{ijk}$, we see that $T^{i'j'k'} a_{j'k'} = v^{i'}$, and that therefore $T^{i'j'k'}$ transforms as a triply covariant tensor, as required.

3 Derivatives (10 points)

Consider a scalar function $f(x^\mu)$ defined on $\mathbb{R}^n$ (an example would be the gravitational potential defined at every point in space in $\mathbb{R}^3$). We can take derivatives of this function, $\partial f / \partial x^\mu$.

- (a) Show that $\partial f / \partial x^\mu$ transforms covariantly. This justifies defining

$$\partial_\mu f \equiv \frac{\partial f}{\partial x^\mu}. \quad (11)$$

with a lower index.

SOLUTION: Under a change of basis, x^μ transforms contravariantly as $x'^\mu = a^\mu_{\nu'} x^{\nu'}$ in the new basis, or equivalently $(a^{-1})^\mu_{\nu'} x^{\nu'} = x^\mu$. Using the chain rule, we have

$$\frac{\partial f}{\partial x'^\mu} = \frac{\partial f}{\partial x^\nu} \frac{\partial x^\nu}{\partial x'^\mu} = (a^{-1})^\nu_\mu \frac{\partial f}{\partial x^\nu}, \quad (12)$$

This therefore transform covariantly, as required.

- (b) Define $R = \sqrt{x^\mu x_\mu}$ on $\mathbb{R}^n$ with a metric $g_{\mu\nu}$ that is constant across $\mathbb{R}^n$ (like the Euclidean or Minkowski metrics). Evaluate $\partial_\mu R$ and $\partial_\mu \partial_\nu R$. (*Hint:*) If this is not straightforward for you, begin by evaluating $\partial_\mu x^\nu$ and $\partial_\mu x_\nu$.

SOLUTION:

First, let's note that

$$\partial_\mu x^\nu = \frac{\partial x^\nu}{\partial x^\mu} = \delta^\nu_\mu, \quad \partial_\mu x_\nu = \frac{\partial x_\nu}{\partial x^\mu} = \frac{\partial}{\partial x^\mu} (g_{\nu\rho} x^\rho) = g_{\nu\rho} \delta^\rho_\mu = g_{\mu\nu}. \quad (13)$$

Then,

$$\partial_\mu R = \frac{1}{2R} \partial_\mu (x^\nu x_\nu) = \frac{1}{2R} (x_\nu \partial_\mu x^\nu + x^\nu \partial_\mu x_\nu) = \frac{1}{2R} (x_\nu \delta^\nu_\mu + x^\nu g_{\mu\nu}) = \frac{1}{2R} (x_\mu + x_\mu) = \frac{x_\mu}{R}. \quad (14)$$

The second derivative is just a bit more algebra:

$$\begin{aligned}\partial_\mu \partial_\nu R &= \partial_\mu \left(\frac{x_\nu}{R} \right) \\ &= \frac{g_{\mu\nu}}{R} - \frac{x_\nu}{R^2} \partial_\mu R \\ &= \frac{1}{R^3} (g_{\mu\nu} R^2 - x_\mu x_\nu)\end{aligned}\tag{15}$$

4 (SG 10.5) Properties of the Levi-Civita Symbol (10 points)

The n -dimensional Levi-Civita symbol is an object $\epsilon_{i_1 i_2 \dots i_n}$ defined by requiring that $\epsilon_{i_1 i_2 \dots i_n}$ be antisymmetric under the swapping of any pair of indices, i.e. $\epsilon_{\dots i \dots j \dots} = -\epsilon_{\dots j \dots i \dots}$ for any two indices. In addition, we set $\epsilon_{12 \dots n} = 1$. Note that it is *not a tensor*, something we will come back to in the lectures.

- (a) Show that if any two indices $i_i = i_j$ for $i, j \in 1, \dots, n$, then $\epsilon_{i_1 i_2 \dots i_n} = 0$.

SOLUTION:

Suppose $i_i = i_j$. Then swapping $i_i \leftrightarrow i_j$ picks up a relative minus sign, but the Levi-Civita symbol remains unchanged. Anything which is the negative of itself must be zero.

- (b) Show that $\epsilon_{123} = \epsilon_{231} = \epsilon_{312}$, but that $\epsilon_{1234} = -\epsilon_{2341} = \epsilon_{3412} = -\epsilon_{4123}$.

SOLUTION:

ϵ_{231} can be obtained from ϵ_{123} by swapping $1 \leftrightarrow 2$ followed by $1 \leftrightarrow 3$. Since this is two swaps, there is no relative sign. Similarly, ϵ_{312} can be obtained from ϵ_{123} by swapping $1 \leftrightarrow 3$ followed by $1 \leftrightarrow 2$. Thus, $\epsilon_{231} = \epsilon_{312} = \epsilon_{123}$.

Similarly, to get ϵ_{2341} from ϵ_{1234} , we need to swap $1 \leftrightarrow 2$, $2 \leftrightarrow 3$, and $3 \leftrightarrow 4$. This is three swaps, and so we get a relative sign of -1 . The other relations can be shown similarly.

- (c) Write the product $\epsilon_{ijk} \epsilon_{i'j'k'}$ as a sum of products of three Kronecker deltas, i.e.

$$\epsilon_{ijk} \epsilon_{i'j'k'} = \delta_i^{i'} \delta_j^{j'} \delta_k^{k'} + 5 \text{ other terms},\tag{16}$$

where you should write out all six terms explicitly.

SOLUTION:

$$\epsilon_{ijk} \epsilon_{i'j'k'} = \delta_i^{i'} \delta_j^{j'} \delta_k^{k'} + \delta_i^{j'} \delta_j^{k'} \delta_k^{i'} + \delta_i^{k'} \delta_j^{i'} \delta_k^{j'} - \delta_i^{i'} \delta_j^{k'} \delta_k^{j'} - \delta_i^{j'} \delta_j^{i'} \delta_k^{k'} - \delta_i^{k'} \delta_j^{j'} \delta_k^{i'}.\tag{17}$$

- (d) Show that $\epsilon_{ijk} \epsilon_{ij'k'} = \delta_j^{j'} \delta_k^{k'} - \delta_{k'}^j \delta_j^{k'}$.

SOLUTION:

$$\begin{aligned}\epsilon_{ijk} \epsilon_{ij'k'} &= \delta_i^{i'} \epsilon_{ijk} \epsilon_{i'j'k'} \\ &= 3\delta_j^{j'} \delta_k^{k'} + 2\delta_j^{k'} \delta_k^{j'} - 3\delta_j^{k'} \delta_k^{j'} - 2\delta_j^{j'} \delta_k^{k'} \\ &= \delta_j^{j'} \delta_k^{k'} - \delta_{k'}^j \delta_j^{k'},\end{aligned}\tag{18}$$

as required.

- (e) For dimension $n = 4$, write out $\epsilon_{ijkl}\epsilon_{i'j'k'l'}$ as a sum of products of Kronecker deltas similar to the one in part (c).

SOLUTION:

We can easily see that in general,

$$\epsilon_{ijkl}\epsilon_{i'j'k'l'} = \text{sgn}(i'j'k'l')\delta_i^{i'}\delta_j^{j'}\delta_k^{k'}\delta_l^{l'} + \text{remaining 23 permutations of } i'j'k'l', \quad (19)$$

where sgn is positive for even permutations of $i'j'k'l'$ and negative for odd permutations. But we can group these permutations as (keeping the)

$$\begin{aligned} \epsilon_{ijkl}\epsilon_{i'j'k'l'} &= \delta_i^{i'} \left(\text{sgn}(i', j'k'l')\delta_j^{j'}\delta_k^{k'}\delta_l^{l'} + \text{remaining 5 permutations of } j'k'l' \right) \\ &\quad + \delta_i^{j'} \left(\text{sgn}(j', i'k'l')\delta_j^{i'}\delta_k^{k'}\delta_l^{l'} + \text{remaining 5 permutations of } i'k'l' \right) \\ &\quad + \delta_i^{k'} \left(\text{sgn}(k', j'i'l')\delta_j^{j'}\delta_k^{i'}\delta_l^{l'} + \text{remaining 5 permutations of } j'i'l' \right) \\ &\quad + \delta_i^{l'} \left(\text{sgn}(l', j'k'i')\delta_j^{j'}\delta_k^{k'}\delta_l^{i'} + \text{remaining 5 permutations of } j'k'i' \right) \end{aligned} \quad (20)$$

$$\begin{aligned} &= \delta_i^{i'} \left(\delta_j^{j'}\delta_k^{k'}\delta_l^{l'} + \text{remaining 5 permutations of } j'k'l' \right) \\ &\quad + (i' \leftrightarrow j') + (i' \leftrightarrow k') + (i' \leftrightarrow l'), \end{aligned} \quad (21)$$

Contracting the full expression with $\delta_i^{i'}$, we see that all the terms in the parentheses of the first four lines will become identical up to an overall sign, since in the second line we have to replace $i' \rightarrow j'$, the third $i' \rightarrow k'$ and the fourth $i' \rightarrow l'$, but the relative sign is negative for the second to fourth line, which is obvious from the structure of the sgn term for the first term in the parentheses. Thus, keeping in mind that $\delta_i^{i'}\delta_i^{i'} = 4$, we have

$$\begin{aligned} \epsilon_{ijkl}\epsilon_{i'j'k'l'} &= (4 - 1 - 1 - 1) \left(\delta_j^{j'}\delta_k^{k'}\delta_l^{l'} + \text{remaining 5 permutations of } j'k'l' \right) \\ &= \delta_i^{i'}\delta_j^{j'}\delta_k^{k'} + \delta_i^{j'}\delta_j^{k'}\delta_k^{i'} + \delta_i^{k'}\delta_j^{i'}\delta_k^{j'} - \delta_i^{i'}\delta_j^{k'}\delta_k^{j'} - \delta_i^{j'}\delta_j^{i'}\delta_k^{k'} - \delta_i^{k'}\delta_j^{j'}\delta_k^{i'}. \end{aligned} \quad (22)$$

5 Index Notation 1 (15 points)

Consider the following tensor in Euclidean $\mathbb{R}^3$:

$$R_{ij} = \delta_{ij} \cos \alpha + n_i n_j (1 - \cos \alpha) - \epsilon_{ijk} n^k \sin \alpha, \quad (23)$$

where n_i is a unit vector. Show that $R_{ji}R^{jk} = \delta_i^k$. R_{ij} represents a rotation about the axis defined by $\hat{n}$, and so therefore must be orthogonal, which is what you're trying to show!

Let's just write everything out:

$$\begin{aligned} R_{ji}R^{jk} &= \delta_{ji}\delta^{jk} \cos^2 \alpha + \delta_{ji}n^j n^k (1 - \cos \alpha) \cos \alpha - \delta_{ji}\epsilon^{jkl} n_l \sin \alpha \cos \alpha \\ &\quad + n_j n_i \delta^{jk} (1 - \cos \alpha) \cos \alpha + n_j n_i n^j n^k (1 - \cos \alpha)^2 - n_j n_i \epsilon^{jkl} n_l \sin \alpha (1 - \cos \alpha) \\ &\quad - \epsilon_{jil} n^l \delta^{jk} \sin \alpha \cos \alpha - \epsilon_{jil} n^l n^j n^k (1 - \cos \alpha) \sin \alpha + \epsilon_{jil} n^l \epsilon^{jkl} n_p \sin^2 \alpha \end{aligned} \quad (24)$$

Notice that I've been very careful to make sure that the free indices agree on both the left-hand side and the right-hand side, changing dummy indices to other letters as necessary.

To make progress, first we note that

$$\epsilon_{jil} n^l \delta^{jk} = -\epsilon_i^{kl} n_l = -\delta_{ij} \epsilon^{jkl} n_l, \quad (25)$$

and so the two terms proportional to $\sin \alpha \cos \alpha$ cancel out. Next, we have

$$\epsilon^{jkl} n_j n_i n_l = 0, \quad (26)$$

since ϵ^{jkl} is antisymmetric under $j \leftrightarrow l$, but $n^l n^j$ is symmetric. A similar argument holds for $\epsilon_{jil} n^l n^j n^k$. Finally, we note that from the previous problem,

$$\epsilon_{jil} n^l \epsilon^{jkp} n_p = (\delta_i^k \delta_l^p - \delta_l^k \delta_i^p) n^l n_p = \delta_i^k - n_i n^k. \quad (27)$$

At this point, we therefore have

$$\begin{aligned} R_{ij} R^{jk} &= \delta_i^k \cos^2 \alpha + 2n_i n^k (1 - \cos \alpha) \cos \alpha + n_i n^k (1 - \cos \alpha)^2 + (\delta_i^k - n_i n^k) \sin^2 \alpha \\ &= \delta_i^k (\cos^2 \alpha + \sin^2 \alpha) + n_i n^k (2 \cos \alpha - 2 \cos^2 \alpha + 1 - 2 \cos \alpha + \cos^2 \alpha - \sin^2 \alpha) \\ &= \delta_i^k. \end{aligned} \quad (28)$$

6 Index Notation 2 (10 points)

Consider an $n \times n$ matrix X with entries X_{ij} , each being an independent variable that can be varied. In many machine learning applications, it is useful to compute the derivative of various quantities involving the matrix X with respect to its entries. A giant compendium of such results can be found [here](#). Let's demystify some of these results using matrix notation. You can assume that matrices act on vectors in Euclidean space, and so index position doesn't matter.

- (a) Show that for an $n \times n$ invertible matrix X ,

$$\frac{\partial X_{kl}^{-1}}{\partial X_{ij}} = -X_{ki}^{-1} X_{jl}^{-1}. \quad (29)$$

(Hint:) start by considering the derivative of the matrix product $X^{-1}X$.

SOLUTION:

Since $X^{-1}X = \mathbb{I}$ the identity matrix, the derivative $\partial(X^{-1}X)/\partial X_{ij} = 0$. But by the product rule, writing $X^{-1}X$ in index notation, we have

$$\begin{aligned} \frac{\partial}{\partial X_{ij}} [X_{kl}^{-1} X_{lm}] &= \frac{\partial X_{kl}^{-1}}{\partial X_{ij}} X_{lm} + X_{kl}^{-1} \frac{\partial X_{lm}}{\partial X_{ij}} \\ &= \frac{\partial X_{kl}^{-1}}{\partial X_{ij}} X_{lm} + X_{kl}^{-1} \delta_{il} \delta_{jm} \\ &= 0. \end{aligned} \quad (30)$$

Therefore,

$$\frac{\partial X_{kl}^{-1}}{\partial X_{ij}} X_{lm} = -X_{ki}^{-1} \delta_{jm}. \quad (31)$$

Multiplying both sides by X_{mn}^{-1} , we find

$$\frac{\partial X_{kl}^{-1}}{\partial X_{ij}} X_{lm} X_{mn}^{-1} = -X_{ki}^{-1} X_{mn}^{-1} \delta_{jm} \implies \frac{\partial X_{kl}^{-1}}{\partial X_{ij}} \delta_{ln} = -X_{ki}^{-1} X_{jn}^{-1}. \quad (32)$$

Relabeling some indices gives

$$\frac{\partial X_{kl}^{-1}}{\partial X_{ij}} = -X_{ki}^{-1} X_{jl}^{-1}, \quad (33)$$

as required.

(b) For a constant matrix A , show that

$$\frac{\partial}{\partial X_{ij}} \text{Tr}(XA) = A_{ji}, \quad (34)$$

where Tr is the trace, and XA is a matrix multiplication.

SOLUTION:

We have

$$\text{Tr}(XA) = X_{km} A_{mk}. \quad (35)$$

Taking the derivative, we find straightforwardly

$$\frac{\partial}{\partial X_{ij}} \text{Tr}(XA) = \delta_{ik} \delta_{jm} A_{mk} = A_{ji}. \quad (36)$$