Appendix A
Linear algebra

The study of the geometry of Lagrangian mechanics requires that one be familiar with basic concepts in abstract linear algebra. The reader is expected to have encountered these concepts before, so this appendix serves as a refresher. We also use our discussion of linear algebra as our "in" to talking about the summation convention in a systematic manner. Since this gets used a lot, the reader may wish to take the opportunity to become familiar with it.

A.1 Vector spaces

We shall suppose the reader to be familiar with the notion of a vector space \( V \) over a field \( K \), particularly the field \( \mathbb{R} \) of real numbers, or the field \( \mathbb{C} \) of complex numbers. On such a vector space, one has defined the notions of vector addition, \( v_1 + v_2 \in V \), between elements \( v_1, v_2 \in V \), and the notion of scalar multiplication, \( av \in V \), for a scalar \( a \in K \) and \( v \in V \). There is a distinguished zero vector \( 0 \in V \) with the property that \( 0 + v = v + 0 = v \) for each \( v \in V \). For the remainder of the section we take \( K \in \{\mathbb{R}, \mathbb{C}\} \).

A subset \( U \subset V \) of a vector space is a subspace if \( U \) is closed under the operations of vector addition and scalar multiplication. If \( V_1 \) and \( V_2 \) are vector spaces, the direct sum of \( V_1 \) and \( V_2 \) is the vector space whose set is \( V_1 \times V_2 \) (the Cartesian product), and with vector addition defined by \( (u_1, u_2) + (v_1, v_2) = (u_1 + v_1, u_2 + v_2) \) and scalar multiplication defined by \( a(u_1, v_2) = (av_1, av_2) \). If \( U_1 \) and \( U_2 \) are subspaces of \( V \) we shall also write \( V = U_1 \oplus U_2 \) if \( U_1 \cap U_2 = \{0\} \) and if every vector \( v \in V \) can be written as \( v = u_1 + u_2 \) for some \( u_1 \in U_1 \) and \( u_2 \in U_2 \).

A collection \( \{v_1, \ldots, v_k\} \) of vectors is linearly independent if the equality
\[
c_1v_1 + \cdots + c_kv_k = 0
\]
holds only when \( c_1 = \cdots = c_k = 0 \). A set of vectors \( \{v_1, \ldots, v_k\} \) generates a vector space \( V \) if every vector \( v \in V \) can be written as
\[
v = c_1v_1 + \cdots + c_kv_k
\]
for some choice of constants \( c_1, \ldots, c_k \in K \). A basis for a vector space \( V \) is a collection of vectors which is linearly independent and which generates \( V \). The number of vectors in a basis we call the dimension of \( V \), and this is readily shown to be independent of choice of basis. A vector space is finite-dimensional if it possesses a basis with a finite number of elements. If \( \{e_1, \ldots, e_n\} \) is a basis for \( V \), we can write
\[
v = v^1e_1 + \cdots + v^ne_n
\]
for some unique choice of \( v^1, \ldots, v^n \in K \), called the components of \( v \) relative to the basis. Here we begin to adopt the convention that components of vectors be written with index up. Let us use this chance to introduce the summation convention we shall employ.

A.1.1 Basic premise of summation convention Whenever one sees a repeated index, one as a subscript and the other as a superscript, summation is implied.

Thus, for example, we have
\[
v^i e_i = \sum_{i=1}^{n} v^i e_i,
\]
as summation over \( i \) is implied.

A map \( A : U \to V \) between is \( k \)-linear if \( A(au) = aA(u) \) and if \( A(u_1 + u_2) = A(u_1) + A(u_2) \) for each \( a \in K \) and \( u, u_1, u_2 \in U \). The linear map \( \text{id}_V : V \to V \) defined by \( \text{id}_V(v) = v \), \( v \in V \), is called the identity map for \( V \). If \( \{f_1, \ldots, f_m\} \) is a basis for \( U \) and \( \{v_1, \ldots, v_n\} \) is a basis for \( V \), for each \( i \in \{1, \ldots, m\} \) we may write
\[
A(f_i) = A^1_i v_1 + \cdots + A^n_i v_n
\]
for some unique choice of constants \( A^1_i, \ldots, A^n_i \in K \). By letting \( a \) run from \( 1 \) to \( m \) we thus define \( mn \) constants \( A^a_i \in K \), \( i = 1, \ldots, n \), \( a = 1, \ldots, m \), which we call the matrix of \( A \) relative to the two bases. If \( u \in U \) is written as
\[
u = u^1 f_1 + \cdots + u^m f_m,
\]
one readily ascertains that
\[
A(u) = \sum_{i=1}^{n} \sum_{a=1}^{m} A^a_i u^a e_i.
\]
Thus the components of \( A(u) \) are written using the summation convention as \( A^a_1 u^1, \ldots, A^n_m u^m \).

Let us say a few more things about our summation convention.

A.1.2 More properties of the summation convention Therefore, in our usual notion of matrix/vector multiplication, this renders the up index for \( A \) the row index, and the down index the column index. Note that we can also compactly write
\[
\sum_{i=1}^{n} \sum_{a=1}^{m} A^a_i u^a e_i = A^a_1 u^1 e_1.
\]

The set of linear maps from a vector space \( U \) to a vector space \( V \) is itself a vector space which we denote \( L(U; V) \). Vector addition in \( L(U; V) \) is given by
\[
(A + B)(u) = A(u) + B(u),
\]
and scalar multiplication is defined by
\[
(aA)(u) = a(A(u)).
\]
Note that what is being defined in these two equations is \( A + B \in L(U; V) \) in the first case, and \( aA \in L(U; V) \) in the second case. One verifies that \( \dim(L(U; V)) = \dim(U) \dim(V) \).
Given a linear map \( A : U \to V \), the kernel of \( A \) is the subspace
\[
\ker(A) = \{ u \in U \mid A(u) = 0 \}
\]
of \( U \), and the image of \( A \) is the subspace
\[
\text{image}(A) = \{ A(u) \mid u \in U \}
\]
of \( V \). The rank of \( A \) is defined to be \( \text{rank}(A) = \dim(\text{image}(A)) \). The rank-nullity formula says that \( \dim(\ker(A)) + \text{rank}(A) = \dim(U) \).

Of special interest are linear maps from a vector space \( V \) to itself: \( A : V \to V \). In this case, an eigenvalue for \( A \) is an element \( \lambda \in K \) with the property that \( A(v) = \lambda v \) for some nonzero vector \( v \), called an eigenvector for \( \lambda \). To compute eigenvalues, one finds the roots of the characteristic polynomial
\[
\det(\lambda I - A) = (\lambda - \lambda_1)^{k_1} \cdots (\lambda - \lambda_k)^{k_k}
\]
det(\( \lambda I - A \)) has degree equal to the dimension of \( V \). If \( K = \mathbb{C} \), this polynomial is guaranteed to have \( \dim(V) \) solutions, but it is possible that some of these will be repeated roots of characteristic polynomial. If
\[
\det(\lambda I - A) = (\lambda - \lambda_1)^k P(\lambda)
\]
then the eigenvalue \( \lambda_0 \) has algebraic multiplicity \( k \). The eigenvectors for an eigenvalue \( \lambda_0 \) are nonzero vectors from the subspace
\[
W_{\lambda_0} = \{ v \in V \mid (A - \lambda_0 I)(v) = 0 \}.
\]
The geometric multiplicity of an eigenvalue \( \lambda_0 \) is \( \dim(W_{\lambda_0}) \). We let \( m_\alpha(\lambda_0) \) denote the algebraic multiplicity and \( m_g(\lambda_0) \) denote the geometric multiplicity of \( \lambda_0 \). It is always the case that \( m_\alpha(\lambda_0) \geq m_g(\lambda_0) \), and both equality and strict inequality can occur.

### A.2 Dual spaces

The notion of a dual space to a vector space \( V \) is extremely important for us. It is also a potential point of confusion, as it seems, for whatever reason, to be a slippery concept.

Given a finite-dimensional vector space \( V \) (let us agree to now restrict to vector spaces over \( \mathbb{R} \)), the dual space to \( V \) is the set \( V^* \) of linear maps from \( V \) to \( \mathbb{R} \). If \( \alpha \in V^* \), we shall alternately write \( \alpha(v) \), \( \alpha \cdot v \), or \( (\alpha; v) \) to denote the image in \( \mathbb{R} \) of \( v \in V \) under \( \alpha \). Note that since \( \dim(\mathbb{R}) = 1 \), \( V^* \) is a vector space having dimension equal to that of \( V \). We shall often call elements on \( V^* \) one-forms.

Let us see how to represent elements in \( V^* \) using a basis for \( V \). Given a basis \( \{ e_1, \ldots, e_n \} \) for \( V \), we define \( n \) elements of \( V^* \), denoted \( e^1, \ldots, e^n \), by \( e^i(e_j) = \delta^i_j, i, j = 1, \ldots, n \), where \( \delta^i_j \) denotes the Kronecker delta
\[
\delta^i_j = \begin{cases} 1, & i = j \\ 0, & \text{otherwise}. \end{cases}
\]

The following result is important, albeit simple.

A.2.1 Proposition: If \( \{ e_1, \ldots, e_n \} \) is a basis for \( V \) then \( \{ e^1, \ldots, e^n \} \) is a basis for \( V^* \), called the dual basis.

Proof: First let us show that the dual vectors \( \{ e^1, \ldots, e^n \} \) are linearly independent. Let \( c_1, \ldots, c_n \in \mathbb{R} \) have the property that
\[
c_1 e^1 + \cdots + c_n e^n = 0.
\]
For each \( j = 1, \ldots, n \) we must therefore have \( c_j e^j(e_j) = c_j \delta^i_j = c_j = 0 \). This implies linear independence. Now let us show that each dual vector \( \alpha \in V^* \) can be expressed as a linear combination of \( \{ e^1, \ldots, e^n \} \).

We claim that \( \alpha = \alpha_1 e^1 + \cdots + \alpha_n e^n \). To check this, it suffices to check that the two one-forms \( \alpha \) and \( \alpha_i e^i \) agree when applied to any of the basis vectors \( \{ e_1, \ldots, e_n \} \). This is obvious since \( e_i = 1 \) if \( i = j \), \( 0 \) otherwise.

### A.3 Bilinear forms

We have multiple opportunities to define mechanical objects that are “quadratic.” Thus the notion of a bilinear form is a useful one in mechanics, although it is unfortunately not normal part of the background of those who study mechanics. However, the ideas are straightforward enough.

We let \( V \) be a finite-dimensional \( \mathbb{R} \)-vector space. A bilinear form on \( V \) is a map \( B : V \times V \to \mathbb{R} \) with the property that for each \( v_0 \in V \) the maps \( v \mapsto B(v, v_0) \) and \( v \mapsto B(v_0, v) \) are linear. Thus \( B \) is “linear in each entry.” A bilinear form \( B \) is symmetric if \( B(v_1, v_2) = B(v_2, v_1) \) for all \( v_1, v_2 \in V \), and skew-symmetric if \( B(v_1, v_2) = -B(v_2, v_1) \) for all \( v_1, v_2 \in V \). If \( \{ e_1, \ldots, e_n \} \) is a basis for \( V \), the matrix for a bilinear form \( B \) in this basis is the collection of \( n^2 \) number \( B_{ij} = B(e_i, e_j) \), \( i, j = 1, \ldots, n \). A form is symmetric if and only if \( B_{ij} = B_{ji} \), \( i, j = 1, \ldots, n \), and skew-symmetric if and only if \( B_{ij} = -B_{ji} \), \( i, j = 1, \ldots, n \).

A.3.1 More properties of the summation convention Note that the indices for the matrix of a bilinear form are both subscripts and superscripts. This should help distinguish bilinear forms from linear maps, since in the latter there is one index up and one index down. If \( B \) is a bilinear form with matrix \( B_{ij} \), and if \( u \) and \( v \) are vectors with components \( u^i, v^j \), \( i = 1, \ldots, n \), then
\[
B(u, v) = B_{ij} u^i v^j.
\]

An important notion attached to a symmetric or skew-symmetric bilinear form is a map from \( V \) to \( V^* \). If \( B \) is a bilinear form which is either symmetric or skew-symmetric, we define
\[
B(u, v) = (u^i v^j) = (v^i u^j) = (v^i e^j(u)) = (e^j(u) v^i) = (u^i v^j) = (u^j v^i).
\]
a map $B^*: V \rightarrow V^*$ by indicating how $B^*(v) \in V^*$ acts on vector in $V$. That is, for $v \in V$ we define $B^*(v) \in V^*$ to be defined by

$$(B^*(v); u) = B(u, v), \quad u \in V.$$  

The rank of $B$ is defined to be $\text{rank}(B) = \dim(\text{image}(B^*))$. $B$ is nondegenerate if $\text{rank}(B) = \dim(V)$. In this case $B^*$ is an isomorphism since $\dim(V) = \dim(V^*)$, and we denote the inverse by $B^T: V^* \rightarrow V$.

A.3.2 More properties of the summation convention If $\{e_1, \ldots, e_n\}$ is a basis for $V$ with dual basis $\{\varepsilon^1, \ldots, \varepsilon^n\}$, then $B^*(v) = B_{ij} \varepsilon^j$. If $B$ is nondegenerate then $B^*(\alpha) = B^*\alpha \varepsilon_1$, where $B^*_{ij}$, $i,j = 1, \ldots, n$, are defined by $B^*_{ij} = \delta_{ij}$.

This statement is the content of Exercise E2.17.

For symmetric bilinear forms, there are additional concepts which will be useful for us. In particular, the following theorem serves as the definition for the index, and this is a useful notion.

A.3.3 Theorem If $B$ is a symmetric bilinear form on a vector space $V$, then there exists a basis $\{e_1, \ldots, e_n\}$ for $V$ so that the matrix for $B$ in this basis is given by

$$
\begin{bmatrix}
1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\
\vdots & \cdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & -1 & \cdots & 0 \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \cdots & \vdots & \cdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 0
\end{bmatrix}
$$

The number of nonzero elements on the diagonal of this matrix is the rank of $B$. The number of $-1$’s along the diagonal is called the index of $B$.

This theorem is proved by the Gram-Schmidt argument. Note that the number of $+1$’s on the diagonal of the matrix in the theorem is given by $\text{rank}(B) - \text{ind}(B)$.

A.4 Inner products

An important special case is when all elements on the diagonal in Theorem A.3.3 have the same sign. If all diagonal elements are $+1$ then we say $B$ is positive-definite and if all diagonal elements are $-1$ then we say $B$ is negative-definite. Clearly, $B$ is positive-definite if and only if $B(v, v) > 0$ whenever $v \neq 0$, and $B$ is negative-definite if and only if $B(v, v) < 0$ whenever $v \neq 0$. A symmetric, positive-definite bilinear form is something with which you are doubtless familiar: it is an inner product.

The most familiar example of an inner product is the standard inner product on $\mathbb{R}^n$. We denote this inner product by $g_{\text{can}}$ and recall that it is defined by

$$g_{\text{can}}(x, y) = \sum_{i=1}^{n} x^i y^i.$$
Let \( v \in V \) and write

\[
v = v^i e_i = \tilde{v}^i f_i.
\]

Using the relation between the basis vectors we have

\[
v^i e_i = v^i Q^j_i f_j = \tilde{v}^j f_j.
\]

Since \( \mathcal{F} \) is linearly independent, this allows us to conclude that the components \( v^i, i = 1, \ldots, n \), and \( \tilde{v}^j, i = 1, \ldots, n \), are related by

\[
\tilde{v}^j = v^i Q^j_i, \quad j = 1, \ldots, n.
\]

Similarly, if \( \alpha \in V^* \), then we write

\[
\alpha = \alpha^i e^i = \tilde{\alpha}^i f^i.
\]

Proceeding as we did for vectors in \( V \), we compute

\[
\alpha^i e^i = \alpha^i P^j_i f^j = \tilde{\alpha}^j f^j.
\]

Since \( \mathcal{F}^* \) is linearly independent we conclude that the components \( \alpha^i, i = 1, \ldots, n \), and \( \tilde{\alpha}^j, i = 1, \ldots, n \), are related by

\[
\tilde{\alpha}^j = \alpha^i P^j_i, \quad j = 1, \ldots, n.
\]

Now let \( A : V \to V \) be a linear map. The matrix of \( A \) in the basis \( \mathcal{E}, \mathcal{A}^j_i, i, j = 1, \ldots, n, \) are defined by

\[
Ae_i = \mathcal{A}^j_i e_j, \quad i = 1, \ldots, n.
\]

Similarly, the matrix of \( A \) in the basis \( \mathcal{F}^*, \tilde{\mathcal{A}}^j_i, i, j = 1, \ldots, n, \) are defined by

\[
Af_i = \tilde{\mathcal{A}}^j_i f_j, \quad i = 1, \ldots, n.
\]

We write

\[
\tilde{\mathcal{A}}^j_i f_i = \mathcal{A}^j_k e_k = P^j_i A^k_i e_k = P^j_i \mathcal{A}^k_j e_k = P^j_i P^k_j Q^j_k f_i, \quad i = 1, \ldots, n.
\]

Therefore, since \( \mathcal{F} \) is linearly independent, we have

\[
\tilde{\mathcal{A}}^j_i = P^j_i \mathcal{A}^k_j Q^k_i, \quad i, \ell = 1, \ldots, n.
\]

Note that this is the usual similarity transformation.

Finally, let us look at a bilinear map \( B : V \times V \to \mathbb{R} \). We let the matrix of \( B \) in the basis \( \mathcal{E} \) be defined by

\[
B_{ij} = B(e_i, e_j), \quad i, j = 1, \ldots, n,
\]

and the matrix of \( B \) in the basis \( \mathcal{F} \) be defined by

\[
\tilde{B}_{ij} = B(f_i, f_j), \quad i, j = 1, \ldots, n.
\]

Note that we have

\[
\tilde{B}_{ij} = B(f_i, f_j) = B(P^k_i e_k, P^l_j e_l) = P^k_i P^l_j B_{kl}, \quad i, j = 1, \ldots, n.
\]

This relates for us the matrices of \( B \) in the two bases.