BIO 5497: LINEAR ALGEBRA

NATHAN BAKER
(BAKER@BIOCHEM.WUSTL.EDU)

Contents

1. Scalars

Definition 1 (Scalar)**.** A scalar is an element of a field (a set of numbers with certain properties: $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}$, etc.).

2. Vectors

Definition 2 (Vector)**.** A vector $\underline{v}$ is an ordered set of scalars from a field and can be written in "row" form

$$\underline{v} = [v_1, v_2, \ldots, v_n] \tag{2.1}$$

or "column" form:

$$\underline{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}. \tag{2.2}$$

Remark. Vectors can be interpreted in many ways; coordinates in space, functions, "binned" sets of data, etc. All of these definitions are useful in different contexts – be flexible.

2.1. Addition.

Definition 3 (Vector addition)**.** The sum of two length-n vectors is given by

$$\begin{aligned} \underline{u} = \underline{v} + \underline{w} &= [v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n] \\ u_i &= v_i + w_i. \end{aligned} \tag{2.3}$$

Definition 4 (Zero)**.** The zero vector (origin) is

$$\underline{0} = [0, 0, \ldots, 0] \tag{2.4}$$

and has the property

$$\underline{0} + \underline{v} = \underline{v}\,. \tag{2.5}$$

Remark. We will encounter lots of different definitions of "0" in this course: scalar, vector, function, matrix, operator, field, distribution, etc.

Definition 5 (Linear combination)**.** Given a set of vectors $\{\underline{v}_i\}$ and a set of scalars $\{\alpha_i\}$, a linear combination is defined as

$$\sum_i \alpha_i \underline{v}_i. \tag{2.6}$$

Definition 6 (Linear dependence)**.** A set of vectors $\{\underline{v}_i\}$ is linearly dependent if there is a non-trivial (not all 0) set of scalars $\{\alpha_i\}$ such that the linear combination is zero:

$$\sum_i \alpha_i \underline{v}_i = 0. \tag{2.7}$$

If a set of vectors is not linearly dependent, it is said to be linearly independent.

Example 1. Suppose you are designing a drug to inhibit a particular enzyme A. This drug has 3 sites for functionalization $i = 1, 2, 3$; each site can be functionalized with 4 different groups $j = 1, 2, 3, 4$. Let each possible functional group control 2 properties of the drug: binding energy to the desired enzyme (ΔG_j^A) and binding energy to a related enzyme ΔG_j^B according to the following scheme:

j	ΔG_j^A	ΔG_j^B
1	-4	0
2	1	3
3	0	1
4	5	-2

Making the incredibly awful assumption that the 3 functionalization sites are independent, design the following drugs:

(1) A drug that binds A more tightly than B.
(2) A drug that binds B more tightly than A.
(3) A drug that binds A and B with the same affinity.

Solution. Each group can be represented by a vector:

$$\underline{v}_j = \left[\Delta G_j^A, \Delta G_j^B\right]. \tag{2.8}$$

We're looking for the right combination of 3 vectors (one for each functionalization site) to give the desired properties:

$$P = \sum_{j=1}^{4} \alpha_{ij}\, \underline{v}_j \tag{2.9}$$

$$\sum_{j=1}^{4} \alpha_{ij} = 3 \tag{2.10}$$

where $\alpha_{ij} = 0, 1, 2, 3$. Clearly, these vectors aren't linearly-independent so we'll most likely encounter multiple (or no) solutions – a situation that is common for most optimization problems. Here are some solutions

(1) $2\alpha_2 + 3\alpha_3 + 11\alpha_4 < 12$
(2) $2\alpha_2 + 3\alpha_3 + 11\alpha_4 > 12$
(3) $2\alpha_2 + 3\alpha_3 + 11\alpha_4 = 12$

subject to the constraints given above. □

2.2. **Multiplication.**

Definition 7 (Scalar-vector multiplication)**.** We've already seen that, given a scalar α and a vector

$$\underline{v} = [v_1, v_2, \dots v_n] \tag{2.11}$$

that

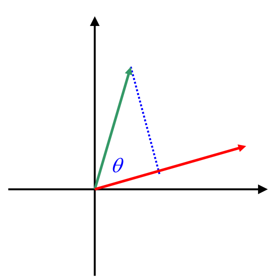
$$\alpha\, \underline{v} = [\alpha v_1, \alpha v_2, \dots \alpha v_n]\,. \tag{2.12}$$

Definition 8 (Inner product)**.** The inner (or scalar) product multiplies two vectors of the same length n according to

$$\begin{aligned} \alpha &= (\,\underline{u}\,,\,\underline{v}\,) = \underline{u}\cdot\underline{w} \\ &= \sum_i^n u_i v_i^*, \end{aligned} \tag{2.13}$$

where the $*$ denotes the complex conjugate.

Remark. Another way to think of the inner product is in terms of the projection of two vectors:



The projection gives the length of one vector in terms the other.

Remark. The inner product is a very powerful concept and is generalized in a number of ways: operators, generalized metric tensors, functions, etc.

Remark. There are several other products, namely the cross products and outer/dyadic products, which are beyond the scope of these lectures. An overview is given in Arfken and Mathworld.

Example 2. One example of inner products occurs in the definition of work:

$$w = \underline{F} \cdot \underline{d} \tag{2.14}$$

due to a displacement $\underline{d}$ against a force $\underline{F}$. Molecular mechanics representations of proteins use this to determine the energy of certain displacements

$$w = \sum_{i}^{n} F_i d_i \tag{2.15}$$

where n is the number of degrees of freedom in the protein. The potential energy change can be obtained from a path integral

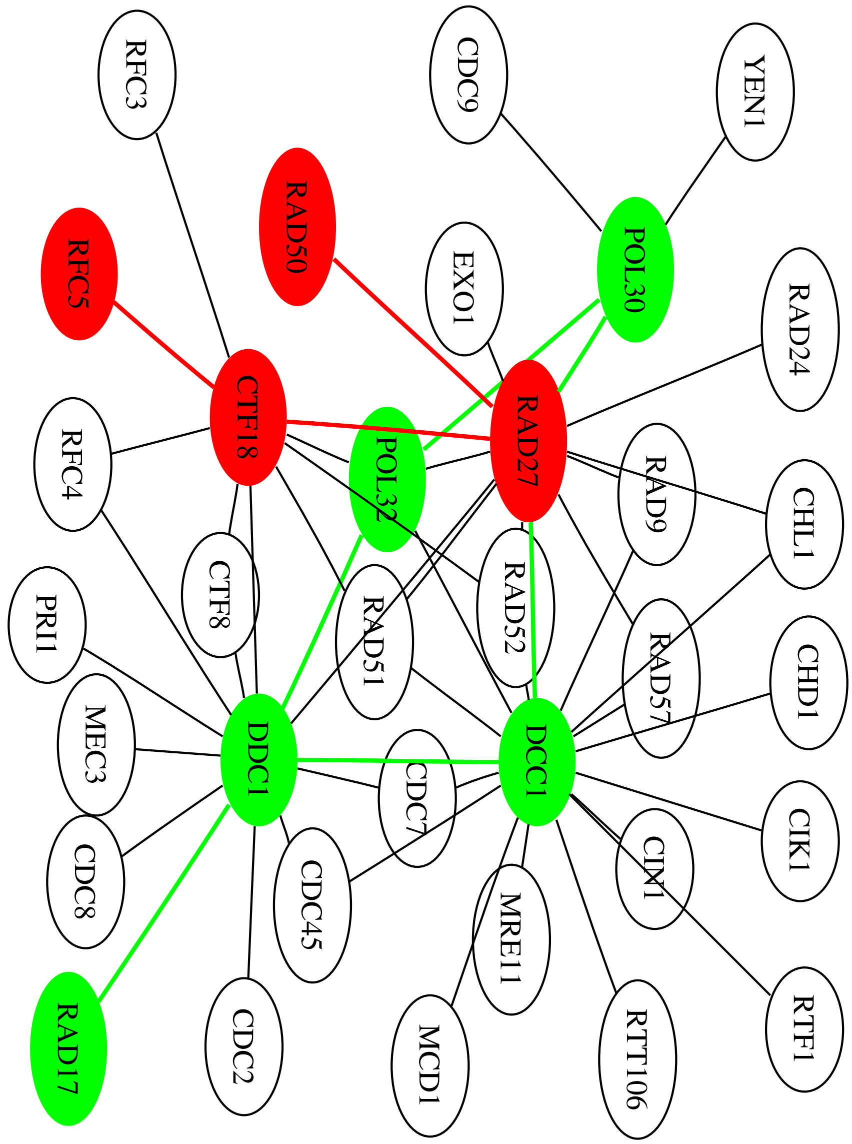
$$\Delta U = \int_{\underline{\ell}} w(\underline{\ell}) d\underline{\ell} = \int_{\underline{\ell}} \underline{F} \cdot d\underline{\ell} \approx \sum_i \underline{F}_i \cdot \Delta\underline{\ell}. \tag{2.16}$$

2.3. **Metrics.**

Definition 9 (Metric)**.** A metric $g(\underline{x}, \underline{y})$ describes the "distance" between vectors $\underline{x}$ and $\underline{y}$ and therefore has the following properties

- $g(\underline{x}, \underline{y}) > 0,\ \underline{x} \neq \underline{y}$
- $g(\underline{x}, \underline{x}) = 0$
- $g(\underline{x}, \underline{y}) = g(\underline{y}, \underline{x})$
- $g(\underline{x}, \underline{y}) + g(\underline{y}, \underline{z}) \geq g(\underline{x}, \underline{z})$

Example 3. Consider a subset of the network of proteins obtained from proteomics work on yeast:



What is a useful metric for this network?

Definition 10 (ℓ^p norm)**.** The generic ℓ^p norm is defined by

$$\| \underline{v} \|_{\ell^p} = \left(\sum_i^n |v_i|^p \right)^{1/p} . \tag{2.17}$$

The most common forms of the ℓ^p norms are $p = 1, 2, \infty$; given below:

- The ℓ^2 (or Euclidean) norm is defined in terms of the inner product:

$$\| \underline{v} \|_{\ell^2} = (\underline{v} \, , \, \underline{v})^{1/2} = \left(\sum_i^n |v_i|^2 \right)^{1/2} . \tag{2.18}$$

  This is the standard Euclidean distance we usually learn about in geometry, etc. It allows us to write the inner product in terms of the angle between two vectors:

$$\underline{u} \cdot \underline{v} = \| \underline{u} \|_{\ell^2} \, \| \underline{v} \|_{\ell^2} \cos \theta; \tag{2.19}$$

or, in higher-dimensional spaces where angle is hard to intuit, it allows us to define the angle. This norm is so ubiquitous, we'll often drop the subscript.

- This norm is based on the absolute value, rather than the square of the vector elements:

$$\| \underline{v} \|_{\ell^1} = \sum_i^n |v_i| . \tag{2.20}$$

- The ℓ^∞ norm sifts the largest element from the vector:

$$\| \underline{v} \|_{\ell^\infty} = \max_i |v_i| . \tag{2.21}$$

Remark. These norms can be extended in a number of ways. First, the L^p norm applies to functions:

$$\|f\|_{L^p(\Omega)} = \left[\int_\Omega |f(x)|^p \, dx \right]^{1/p} . \tag{2.22}$$

Next, the inner products and norms are often defined with respect to a weighting function:

$$(u, v)_g = \int u^*(x) v(x) g(x) dx. \tag{2.23}$$

or an operator

$$(u, v)_A = \int u^*(x) \left[A v(x) \right] dx. \tag{2.24}$$

2.4. **Orthogonality.**

Definition 11 (Orthogonality)**.** Two non-zero vectors $\underline{v}$ and $\underline{w}$ are orthogonal if

$$\underline{v} \cdot \underline{w} = 0 \text{ for } \underline{v} \neq \underline{w} . \tag{2.25}$$

Definition 12 (Orthonormality)**.** Two non-zero vectors $\underline{v}$ and $\underline{w}$ are orthogonal if

$$\underline{v} \cdot \underline{w} = \begin{cases} 1 & \text{for } \underline{v} = \underline{w} \\ 0 & \text{otherwise} . \end{cases} \tag{2.26}$$

Remark. Orthogonal sets of vectors are much easier to deal with. For example, let $\{ \underline{\phi}_i \}$ be a set of vectors and let

$$\underline{u} = \sum_i \alpha_i \underline{\phi}_i \tag{2.27}$$

$$\underline{v} = \sum_i \beta_i \underline{\phi}_i \tag{2.28}$$

be two linear combinations of these vectors. Operations which require the evaluation of $\underline{u} \cdot \underline{v}$ are much simpler with orthogonal vectors:

$$\underline{u} \cdot \underline{v} = \sum_i \alpha_i \beta_i \phi_i^2 ; \tag{2.29}$$

no cross-terms are needed.

Example 4. At what angle are two vectors orthogonal?

Solution.

$$0 = \underline{u} \cdot \underline{v} = \|u\| \, \|v\| \cos\theta$$
$$0 = \cos\theta$$
$$\theta = \frac{\pi + n}{2} \text{ for } n \in \mathbb{Z}. \tag{2.30}$$

□

2.5. **Vector spaces.**

Definition 13. A vector space $V(F)$ is a set of vectors over the field F. This set must be closed under addition

$$\underline{w} = (\underline{u} + \underline{v}) \in V(F) \text{ if } \underline{u}, \underline{v} \in V(F) \tag{2.31}$$

and finite scalar multiplication:

$$\underline{w} = (\alpha \underline{u}) \in V(F) \text{ if } \underline{u}, \in V(F), |\alpha| < \infty, \alpha \in F. \tag{2.32}$$

Remark. Vector spaces are defined in a number of ways, including directly from the field:

$$\mathbb{R}^n = \{[v_1, \ldots, v_n] : v_i \in \mathbb{R}\} \tag{2.33}$$

or subsets of the field obeying certain conditions

$$V(\mathbb{R}) = \{\| \underline{v} \|_{\ell^2} < \infty : \underline{v} \in \mathbb{R}^n\}. \tag{2.34}$$

Definition 14 (Span)**.** Given a set of vectors $V = \{\underline{v}_i\}$, the span is the space of vectors arising from linear combinations of the set elements:

$$\text{span}(V) = \left\{ \sum_i \alpha_i \underline{v}_i : \underline{v}_i \in V, \alpha_i \in \mathbb{R} \right\}. \tag{2.35}$$

Remark. The span is the set of all vectors which can be represented by the set V.

Definition 15 (Basis)**.** The basis of a vector space V is a set W of linearly-independent vectors which span the vector space:

$$\text{span}(W) = V. \tag{2.36}$$

Remark. As with everything else, bases can be developed for discrete vectors and functions. In fact, much of the remainder of the course will deal directly and indirectly with different functional bases for solving problems.

Example 5. One example where bases are particularly practical is in data management with large sets of mutagenesis, etc. data.

Solution. For example, assume you have a protein that has undergone extensive alanine scanning mutagenesis – but not in a terribly systematic fashion. In particular, you have a set of phenotypes (scalars) associated with vectors denoting the position(s) at which the protein was modified. In order to best represent the data, you should find the basis which:

- represents each position once, or
- represents the positions with greatest change, or...

□

Definition 16 (Dimension)**.** The dimension of a space is the size of its basis set – *not* the number of elements in its vectors.

Remark. So, hopefully you're convinced that we need orthonormal bases of our spaces... how do we construct them?

Algorithm 1 (Gram-Schmidt)**.** Let $\{u_i\}$ be a set of non-orthogonal linearly-independent vectors on the space V. The following procedure constructs new sets of orthogonal $\{v_i\}$ and orthonormal $\{w_i\}$ vectors with respect to arbitrary inner products (and corresponding norms).

- Let

$$\begin{aligned} v_0 &= u_0, \\ w_0 &= \frac{u_0}{\|u_0\|} \end{aligned}$$

- For $i = 1, 2, \ldots$

$$\begin{aligned} v_i &= u_i + \sum_{j=0}^{i-1} a_{ij} w_j \\ a_{ij} &= -(u_i, w_j) \\ w_i &= \frac{v_i}{\|v_i\|} \end{aligned}$$

Remark. This algorithm assumes the new basis can be chosen as a linear combination of the old basis and determines the coefficients of the expansion to ensure orthogonality.

Example 6. Suppose you are trying to fit a curve over $[0, 1]$ in terms of the polynomials up to cubic order $V = \{1, x, x^2, x^3\}$. The non-orthogonal basis is very tedious to work with – orthogonalize this with respect to the ℓ^2 inner product.

Solution. Apply the Gram-Schmidt algorithm:

$i = 0$

$$\begin{aligned} v_0 &= 1 \\ w_0 &= 1. \end{aligned}$$

$i = 1$

$$\begin{aligned} a_{10} &= -\int_0^1 x dx = -\frac{1}{2} \\ v_1 &= u_1 + a_{10} w_0 = x - \frac{1}{2} \\ w_1 &= 2\sqrt{3}\left(x - \frac{1}{2}\right) \end{aligned}$$

$i = 2$

$$\begin{aligned}
a_{20} &= -\int_0^1 x^2 dx = -\frac{1}{2} \\
a_{21} &= -\int_0^1 x^2 \cdot 2\sqrt{3}\left(x - \frac{1}{2}\right) dx = -\frac{1}{2\sqrt{3}} \\
v_2 &= x^2 - x + \frac{1}{6} \\
w_2 &= 6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right)
\end{aligned}$$

$i = 3$

$$\begin{aligned}
a_{30} &= -\int_0^1 x^3 dx = -\frac{1}{4} \\
a_{31} &= -\int_0^1 x^3 \cdot 2\sqrt{3}\left(x - \frac{1}{2}\right) dx = -\frac{3\sqrt{3}}{20} \\
a_{32} &= -\int_0^1 x^3 \cdot 6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right) dx = -\frac{1}{4\sqrt{5}} \\
v_3 &= -\frac{1}{4} - \frac{9}{10}\left(x - \frac{1}{2}\right) + x^3 - \frac{3}{2}\left(x^2 - x + \frac{1}{6}\right) \\
w_3 &= 20\sqrt{7}\left[-\frac{1}{4} - \frac{9}{10}\left(x - \frac{1}{2}\right) + x^3 - \frac{3}{2}\left(x^2 - x + \frac{1}{6}\right)\right]
\end{aligned}$$

□

Theorem 1 (Parseval). *Given an orthonormal basis $\{\underline{v}_i\}$ for a space V, then any $\underline{u} \in V$ can be represented as a linear combination:*

$$\underline{u} = \sum_i (\underline{u}, \underline{v}_i)\, \underline{v}_i. \tag{2.37}$$

Proof. Given that $\underline{u} \in V$ implies

$$\underline{u} = \sum_i \alpha_i \underline{v}_i \tag{2.38}$$

for a basis $\{\underline{v}_i\}$. We can find α_i by constructing a set of linear equations:

$$(\underline{v}_j, \underline{u}) = \sum_i \alpha_i (\underline{v}_j, \underline{v}_i). \tag{2.39}$$

However the orthonormality of the basis implies

$$(\underline{v}_j, \underline{v}_i) = \delta_{ij} \tag{2.40}$$

(where δ_{ij} is the Kronecker delta symbol) and therefore

$$(\underline{v}_j, \underline{u}) = \alpha_i. \tag{2.41}$$

□

Remark. This also works as a method for approximating a function by projecting it onto a basis.

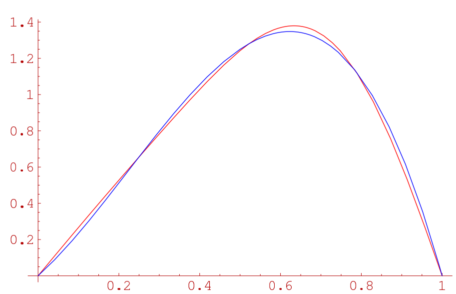
Example 7. The sine basis $2\sin(\pi n x)$ $n \in \mathbb{Z}_+$ is orthonormal on the interval $(0, 1)$. Represent $p(x) = 2x + 4x^2 - 6x^3$ in terms of this basis.

Solution. Using Parseval's theorem, we get

$$p(x) = 2\sum_n \alpha_n \sin(\pi n x)$$

$$\alpha_n = 2\int_0^1 \sin(\pi n x)\, p(x) dx = -\frac{72}{n^3\pi^3}. \tag{2.42}$$

The result is shown in below for a 3-term sine-series (red) and the original function (blue):



□

3. Matrices

Definition 17 (Linear operator)**.** A linear operator is a transformation which maps a vector from a vector space to a vector space:

$$A : V(F) \mapsto W(F) \tag{3.1}$$

in a linear fashion; e.g.,

$$A\left(\alpha\,\underline{v}\,+\beta\,\underline{w}\,\right)=\alpha A\,\underline{v}\,+\beta A\,\underline{w}\,, \tag{3.2}$$

where $\underline{\underline{A}}$ is a linear operator, α,β are scalars, and $\underline{v}\,,\,\underline{w}$ are vectors.

Remark. We're being very general on purpose; linear operators include derivatives, integrals, rotations, scaling, convolutions, etc.

Definition 18 (Matrix)**.** Suppose we have a linear operator $A:U\mapsto V$ and bases $\operatorname{span}\{\,\underline{u}_{\,i}\}=U$ and $\operatorname{span}\{\,\underline{v}_{\,i}\}=V$. Given this information, we can write the result of operating upon $\underline{\phi}\,=\sum_i\,\underline{u}_{\,i}$ with A in terms of a linear combination of $\underline{v}_{\,i}$:

$$A\,\underline{u}_{\,i}=\sum_j a_{ij}\,\underline{v}_{\,j}. \tag{3.3}$$

The coefficients a_{ij} represent a matrix:

$$\underline{\underline{A}}\,=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}. \end{bmatrix}; \tag{3.4}$$

i.e., the discretization of the operator in a particular vector space basis. This is more obvious when the bases are orthonormal, then we have

$$\alpha_{ij}=\left(\,\underline{v}_{\,j},A\,\underline{u}_{\,i}\right). \tag{3.5}$$

Example 8. What is the matrix representation of the Laplacian $-\nabla^2$ in the basis $\{\sin\pi nx\}$ on the interval $(0,1)$?

Solution.

$$-\int_0^1\sin\left(\pi mx\right)\nabla^2\sin\left(\pi nx\right)dx=\frac{n\pi}{2}\delta_{nm}, \tag{3.6}$$

a diagonal matrix. □

Definition 19 (Matrix-matrix addition)**.** The sum of matrix $\underline{\underline{A}}$ with elements $\{a_{ij}\}$ and matrix $\underline{\underline{B}}$ with elements $\{b_{ij}\}$ is a matrix $\underline{\underline{C}}$ with elements

$$c_{ij}=a_{ij}+b_{ij}. \tag{3.7}$$

Definition 20 (Matrix-vector multiplication)**.** Given a matrix $\underline{\underline{A}}$ with elements $\{a_{ij}\}$ and vector $\underline{v}$ with elements $\{v_i\}$, their product is a vector $\underline{u}\,=\,\underline{\underline{A}}\,\underline{v}$ with elements:

$$u_i=\sum_j a_{ij}v_j, \tag{3.8}$$

a sum over the columns of $\underline{\underline{A}}$.

Example 9. What is the matrix representation of rotation by θ in the xy-plane?

Solution. Basis: x and y axes; sign of angle determines direction of rotation.

$$\begin{aligned} z' &= e^{-\imath\theta}z\\ &= \left(\cos\theta-\imath\sin\theta\right)\left(x+\imath y\right)\\ &= \left(x\cos\theta+y\sin\theta\right)+\imath\left(x\sin\theta-y\cos\theta\right) \end{aligned} \tag{3.9}$$

or

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{3.10}$$

□

Definition 21 (Matrix-matrix multiplication)**.** The sum of matrix $\underline{\underline{A}}$ with elements $\{a_{ij}\}$ and matrix $\underline{\underline{B}}$ with elements $\{b_{ij}\}$ is a matrix $\underline{\underline{C}}$ with elements

$$c_{ij} = \sum_k a_{ik} b_{kj}, \tag{3.11}$$

where obviously the number of columns in $\underline{\underline{A}}$ must be the same as the number of rows in $\underline{\underline{B}}$.

Example 10. What is the matrix representation of rotation by θ in the xy-plane followed by rotation of by ϕ in the yz-plane?

Solution. The first rotation (by itself) is

$$\underline{\underline{A}} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.12}$$

The second rotation (by itself) is

$$\underline{\underline{B}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix}. \tag{3.13}$$

Therefore, the application of these two operations (in order) is

$$\underline{\underline{C}} = \underline{\underline{B}}\,\underline{\underline{A}} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\cos\phi\sin\theta & \cos\phi\cos\theta & \sin\phi \\ \sin\phi\sin\theta & -\cos\theta\sin\phi & \cos\phi \end{bmatrix}. \tag{3.14}$$

□

Lemma 1 (Non-commutivity)**.** *In general, for two matrices* $\underline{\underline{A}}$ *and* $\underline{\underline{B}}$,

$$\underline{\underline{A}}\,\underline{\underline{B}} \neq \underline{\underline{B}}\,\underline{\underline{A}}. \tag{3.15}$$

Definition 22 (Transpose)**.** The transpose $\underline{\underline{A}}^T$ of a matrix $\underline{\underline{A}}$ swaps columns and rows:

$$\left(\underline{\underline{A}}^T\right)_{ij} = \left(\underline{\underline{A}}\right)_{ji}. \tag{3.16}$$

Definition 23 (Symmetric matrix)**.** A symmetric matrix obeys the equality $\underline{\underline{A}}^T = \underline{\underline{A}}$.

Lemma 2.

$$\left(\underline{\underline{A}}\,\underline{\underline{B}}\right)^T = \underline{\underline{B}}^T \underline{\underline{A}}^T. \tag{3.17}$$

Example 11. What is the transpose of the xy-plane rotation matrix in the previous example? What does it mean?

Solution. Our matrix is

$$\underline{\underline{A}} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.18}$$

It's transpose is

$$\underline{\underline{A}}^T = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.19}$$

This implies

$$\begin{aligned} x' &= x\cos\theta - y\sin\theta \\ y' &= x\sin\theta + y\cos\theta \end{aligned}$$

(3.20)

or, in complex coordinates,

$$\begin{aligned} z' &= (x\cos\theta - y\sin\theta) + \imath\,(x\sin\theta + y\cos\theta) \\ &= (\cos\theta - \imath\sin\theta)\,(x + \imath y) \\ &= e^{\imath\theta} z. \end{aligned} \tag{3.21}$$

This is an inversion of the previous rotation; i.e., rotation in the opposite direction. □

3.1. **Inversion.**

Remark. We mentioned that matrices represent transforms between two vector spaces. What are the size of these spaces? When can this transformation be "undone"?

Definition 24 (Null space and nullity)**.** The null space (or kernel) of an operator $\underline{\underline{A}}$ is the space of vectors $\{\,\underline{u}_i : \underline{u}_i \neq \underline{0}\,\}$ such that

$$\underline{\underline{A}}\,\underline{u}_i = \underline{0}\,. \tag{3.22}$$

The dimension of the null space is the nullity of $\underline{\underline{A}}$.

Remark. Who cares? If more than one vector gives $\underline{0}$ as a result, then there is not a unique way to undo the transformation...

Lemma 3. *A matrix with a non-zero nullity has linearly-dependent rows and/or columns.*

Proof. We have two sets of whose linear combination gives the same result... □

Definition 25 (Range and rank)**.** The range of the operator $\underline{\underline{A}}$ is the space spanned by

$$\underline{w}_i = \underline{\underline{A}}\,\underline{v}_i, \tag{3.23}$$

where $\{\,\underline{v}_i\}$ is a set of vectors spanning the domain of the operator.

Definition 26 (Dimension)**.** The dimension of an operator is the sum of the rank and nullity.

Remark. Given these concepts, it is now possible to define the operation which "undoes" a transformation: the inverse.

Definition 27 (Identity)**.** The identity operator $\underline{\underline{I}}$ leaves vectors unchanged:

$$\underline{\underline{I}}\,\underline{v} = \underline{v}\,. \tag{3.24}$$

and has elements $I_{ij} = \delta_{ij}$.

Definition 28 (Inverse)**.** The inverse $\underline{\underline{A}}^{-1}$ of an operator $\underline{\underline{A}}$ has the property

$$\underline{\underline{A}}^{-1}\underline{\underline{A}} = \underline{\underline{I}}\,. \tag{3.25}$$

Theorem 2 (Existence of inverse)**.** *If the operator $\underline{\underline{A}}$ has zero nullity; i.e.,*

$$\underline{\underline{A}}\,\underline{v} = \underline{0} \;\text{ iff }\; \underline{v} = 0 \tag{3.26}$$

then the inverse $\underline{\underline{A}}^{-1}$ exists and $\underline{\underline{A}}$ is said to be invertible.

Proof. See Halmos. □

Lemma 4. *If $\underline{\underline{A}}$ and $\underline{\underline{B}}$ are invertible, then*

$$\left(\underline{\underline{B}}\,\underline{\underline{A}}\right)^{-1} = \underline{\underline{A}}^{-1}\underline{\underline{B}}^{-1}. \tag{3.27}$$

Proof. How do we undo the operation $\underline{\underline{B}}\,\underline{\underline{A}}$?

$$\begin{aligned}
\underline{\underline{B}}\,\underline{\underline{A}}\,\underline{x} &= \underline{y} \\
\underline{\underline{B}}^{-1}\underline{\underline{B}}\,\underline{\underline{A}}\,\underline{x} &= \underline{\underline{B}}^{-1}\underline{y} \\
\underline{\underline{A}}\,\underline{x} &= \underline{\underline{B}}^{-1}\underline{y} \\
\underline{\underline{A}}^{-1}\underline{\underline{A}}\,\underline{x} &= \underline{\underline{A}}^{-1}\underline{\underline{B}}^{-1}\underline{y} \\
\underline{x} &= \underline{\underline{A}}^{-1}\underline{\underline{B}}^{-1}\underline{y}\,. \qquad (3.28)
\end{aligned}$$

□

Lemma 5. *If $\underline{\underline{A}}$ is invertible, then $\underline{\underline{A}}^{-1}$ is invertible and*

$$\left(\underline{\underline{A}}^{-1}\right)^{-1} = \underline{\underline{A}}\,. \tag{3.29}$$

Example 12. What is the inverse of rotation in the xy-plane?

Solution. We want to find $\underline{\underline{B}}$ such that

$$\begin{aligned}
\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
\begin{bmatrix} b_{11}\cos\theta - b_{12}\sin\theta & b_{12}\cos\theta + b_{11}\sin\theta \\ b_{21}\cos\theta - b_{22}\sin\theta & b_{22}\cos\theta + b_{21}\sin\theta \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\end{aligned}$$

This gives

$$\begin{aligned}
b_{11}\cos\theta - b_{12}\sin\theta &= 1 \\
b_{12}\cos\theta + b_{11}\sin\theta &= 0 \\
b_{21}\cos\theta - b_{22}\sin\theta &= 0 \\
b_{22}\cos\theta + b_{21}\sin\theta &= 1.
\end{aligned}$$

We can solve these equations to find

$$\underline{\underline{B}} = \underline{\underline{A}}^{-1} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \underline{\underline{A}}^{T}. \tag{3.30}$$

□

Definition 29 (Orthogonal operator)**.** An orthogonal operator is one where

$$\underline{\underline{A}}^{-1} = \underline{\underline{A}}^{T}, \tag{3.31}$$

i.e., all the rows are orthonormal.

3.2. Eigenanalysis.

Remark. There are certain bases which are particular to a given matrix or operator. These bases describe the characteristic modes of the operator and are called eigenvectors. Examples include: normal modes of a protein, resonant frequencies of a harmonic system, principal axes of a solid object, principal components of a probability distribution, etc....

Definition 30 (Eigenvector and eigenvalue)**.** Given a matrix $\underline{\underline{A}}$, the equation

$$\underline{\underline{A}}\,\underline{x} = \lambda\,\underline{x} \tag{3.32}$$

determines pairs of scalars λ called eigenvalues and vectors $\underline{x}$ called eigenvectors.

Theorem 3. *If a matrix has a zero eigenvalue, then it is singular (non-invertible).*

Proof. This follows from the definition of eigenvalues (above) and Thm. 2 (p. 12). □

Definition 31 (Adjoint)**.** The adjoint $\underline{\underline{A}}^{\dagger}$ of a matrix $\underline{\underline{A}}$ has elements

$$\left(\underline{\underline{A}}^{\dagger}\right)_{ij} = \left(\underline{\underline{A}}\right)^{*}_{ji}, \tag{3.33}$$

where $*$ denotes the complex conjugate.

Definition 32 (Self-adjoint matrices)**.** A self-adjoint matrix $\underline{\underline{A}}$ is square and has real or complex elements which satisfy

$$\underline{\underline{A}}^{\dagger} = \underline{\underline{A}}. \tag{3.34}$$

For complex matrices, self-adjoint operators are also called Hermitian. For real matrices, they are called symmetric.

Theorem 4 (Eigenvalues of self-adjoint matrices)**.** *The eigenvalues of a self-adjoint matrix are real.*

Proof. See Arfken. □

Theorem 5 (Eigenvectors of self-adjoint matrices)**.** *The eigenvectors of a self-adjoint matrix are orthogonal.*

Proof. We will prove this for the real symmetric case; see Arfken for the complex case. Suppose we have two eigenpairs

$$\begin{aligned} \underline{\underline{A}}\,\underline{x}_i &= \lambda_i\,\underline{x}_i \\ \underline{\underline{A}}\,\underline{x}_j &= \lambda_j\,\underline{x}_j \end{aligned}$$

and we multiply each eigenvector by the other:

$$\begin{aligned} \left(\underline{x}_j, \underline{\underline{A}}\,\underline{x}_i\right) &= \lambda_i\left(\underline{x}_j, \underline{x}_i\right) \\ \left(\underline{x}_i, \underline{\underline{A}}\,\underline{x}_j\right) &= \lambda_j\left(\underline{x}_j, \underline{x}_j\right). \end{aligned}$$

However, the two left-hand sides are related by the adjoint operator:

$$\left(\underline{x}_i, \underline{\underline{A}}\,\underline{x}_j\right) = \left(\underline{x}_j, \underline{\underline{A}}^\dagger \underline{x}_i\right) \tag{3.35}$$

and, since our operator is self adjoint,

$$\left(\underline{x}_i, \underline{\underline{A}}\,\underline{x}_j\right) = \left(\underline{x}_j, \underline{\underline{A}}\,\underline{x}_i\right). \tag{3.36}$$

This implies

$$\lambda_i \left(\underline{x}_j, \underline{x}_i\right) = \lambda_j \left(\underline{x}_j, \underline{x}_j\right).$$

If $\lambda_i \neq \lambda_j$ (non-degenerate eigenvalues) and $\underline{x}_i, \underline{x}_i \neq \underline{0}$ (non-singular), then we must have

$$\left(\underline{x}_j, \underline{x}_j\right) = 0. \tag{3.37}$$

If the eigenvalues are degenerate, the degeneracy can be broken by small perturbations. □

Example 13. The analysis of vibrational normal modes in molecules is a useful first-order method for determining their IR absorption spectra. Consider the linear molecule O=C=S (carbonyl sulfide). Suppose the bonds in this linear molecule are harmonic; i.e., the potential energy of the system is

$$U(x_O, x_C, x_S) = \frac{1}{2}\left[k_{OC}\left(x_O - x_C\right)^2 + k_{CS}\left(x_C - x_S\right)^2\right]. \tag{3.38}$$

Denoting the masses of the atoms as m_O, m_C, m_S, what is the eigenvalue problem which gives the characteristic frequencies of vibration in this system?

Solution. We want to solve Newton's equations $\underline{F} = m\,\underline{a}$:

$$-\frac{\partial U}{\partial x_i} = m_i \ddot{x}_i. \tag{3.39}$$

which gives

$$\begin{aligned} m_O \ddot{x}_O &= -k_{OC}\left(x_O - x_C\right) \\ m_C \ddot{x}_C &= -k_{OC}\left(x_C - x_O\right) - k_{CS}\left(x_C - x_S\right) \\ m_S \ddot{x}_S &= -k_{CS}\left(x_S - x_C\right). \end{aligned}$$

Assuming vibrations of the form $x_i(t) = x_i(\omega) e^{\imath\omega t}$ gives

$$\begin{aligned} \omega^2 x_O(\omega) &= \frac{k_{OC}}{m_O}\left(x_O(\omega) - x_C(\omega)\right) \\ \omega^2 x_C(\omega) &= \frac{k_{OC}}{m_C}\left(x_C(\omega) - x_O(\omega)\right) + \frac{k_{CS}}{m_C}\left(x_C(\omega) - x_S(\omega)\right) \\ \omega^2 x_S(\omega) &= \frac{k_{CS}}{m_S}\left(x_S(\omega) - x_C(\omega)\right), \end{aligned}$$

where $e^{\imath\omega t}$ can be factored out of the problem. This is an eigenvalue problem of the form

$$\begin{bmatrix} \frac{k_{OC}}{m_O} & -\frac{k_{OC}}{m_O} & 0 \\ -\frac{k_{OC}}{m_C} & \frac{k_{OC}}{m_O} + \frac{k_{CS}}{m_C} & -\frac{k_{CS}}{m_C} \\ 0 & -\frac{k_{CS}}{m_S} & \frac{k_{CS}}{m_S} \end{bmatrix} \begin{bmatrix} x_O(\omega) \\ x_C(\omega) \\ x_S(\omega) \end{bmatrix} = \omega^2 \begin{bmatrix} x_O(\omega) \\ x_C(\omega) \\ x_S(\omega) \end{bmatrix}. \tag{3.40}$$

□

Remark. Normal modes can describe larger molecules, too...

Theorem 6. *The eigenvectors of a matrix form a basis which diagonalizes that matrix.*

Proof. Let $\underline{\underline{A}}$ be the (SPD) matrix under consideration with eigenvectors and eigenvalues:

$$\underline{\underline{A}}\,\underline{x}_i = \lambda_i \underline{x}_i. \tag{3.41}$$

Let $\underline{\underline{W}}$ be the matrix with the eigenvectors $\underline{x}_i$ as columns:

$$\underline{\underline{W}} = \begin{bmatrix} (\underline{x}_1)_1 & (\underline{x}_2)_1 & \cdots & (\underline{x}_N)_1 \\ (\underline{x}_1)_2 & (\underline{x}_2)_2 & \cdots & (\underline{x}_N)_2 \\ \vdots & \vdots & \ddots & \vdots \\ (\underline{x}_1)_N & (\underline{x}_2)_N & \cdots & (\underline{x}_N)_N \end{bmatrix}. \tag{3.42}$$

This implies

$$\underline{\underline{A}}\,\underline{\underline{W}} = \underline{\underline{W}}\,\underline{\underline{\Lambda}} \tag{3.43}$$

where $\underline{\underline{\Lambda}}$ is a diagonal matrix with elements $\left(\underline{\underline{\Lambda}}\right)_{ij} = \lambda_i \delta_{ij}$. Assuming the eigenvectors are orthonormal (possibly by construction), then

$$\underline{\underline{W}}^{\dagger}\underline{\underline{W}} = \underline{\underline{I}} \tag{3.44}$$

which implies

$$\underline{\underline{W}}^{\dagger}\underline{\underline{A}}\,\underline{\underline{W}} = \underline{\underline{\Lambda}}. \tag{3.45}$$

□

Example 14. Solve the hopelessly abstract (but highly relevant) equation

$$\underline{\underline{A}}\,\underline{y} = \underline{b} \tag{3.46}$$

using the eigenvalues of $\underline{\underline{A}}$.

Solution.

$$\begin{aligned} \underline{\underline{A}}\,\underline{y} &= \underline{b} \\ \underline{\underline{A}}\,\underline{\underline{W}}\,\underline{\underline{W}}^{\dagger}\underline{y} &= \underline{b} \\ \underline{\underline{W}}\,\underline{\underline{\Lambda}}\,\underline{\underline{W}}^{\dagger}\underline{y} &= \underline{b} \\ \underline{\underline{\Lambda}}\,\underline{\underline{W}}^{\dagger}\underline{y} &= \underline{\underline{W}}^{\dagger}\underline{b} \\ \underline{\underline{W}}^{\dagger}\underline{y} &= \underline{\underline{\Lambda}}^{-1}\underline{\underline{W}}^{\dagger}\underline{b} \\ \underline{y} &= \underline{\underline{W}}\,\underline{\underline{\Lambda}}^{-1}\underline{\underline{W}}^{\dagger}\underline{b}. \end{aligned} \tag{3.47}$$

Or, in other words, $\underline{\underline{A}}^{-1} = \underline{\underline{W}}\,\underline{\underline{\Lambda}}^{-1}\underline{\underline{W}}^{\dagger}$. □

Example 15. Principal component analysis is used for datasets which describe covariation

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{3.48}$$

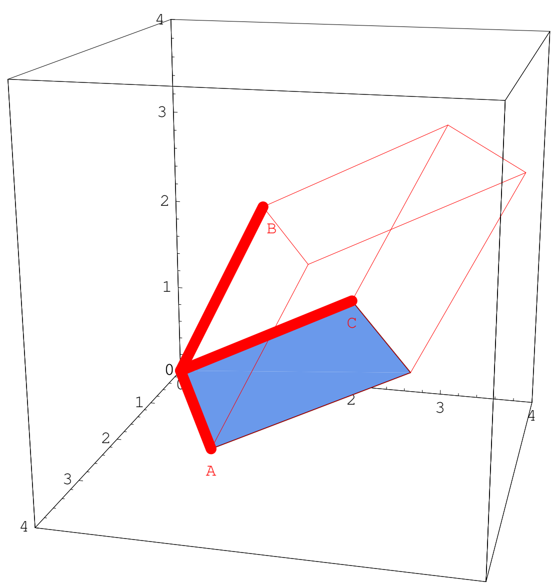
The eigenvectors of this matrix describe the directions of maximum variation, with the largest eigenvalues corresponding to the largest deviations. Principal component analysis is also referred to as "quasiharmonic analysis". Why?

Solution. Suppose the observed covariation is due to a dynamics trajectory or other simulation of a system. By considering only second-order moments, it is implicitly assuming harmonic Hamiltonian for this system with a Hessian related to the inverse of the covariance matrix. □

3.3. Determinants.

Remark. Determinants measure *volume.*

Definition 33 ($\mathbb{R}^3$ volumes as triple scalar products)**.** Consider the parallelepiped written in terms of the vectors $\underline{A}$, $\underline{B}$, $\underline{C}$:



Taking the cross product of the base vectors gives a perpendicular vector whose length is proportional to the area of the base

$$\underline{F} = \underline{A} \times \underline{C} \tag{3.49}$$

$$F = AC\sin\theta. \tag{3.50}$$

The dot product of this new vector with $\underline{A}$ is the "base times the height" formula for the volume:

$$\begin{aligned} V &= (\underline{A} \times \underline{C}) \cdot \underline{B} \\ &= A_x (B_y C_z - B_z C_y) + A_y (B_z C_x - B_x C_z) + A_z (B_x C_y - B_y C_x) \end{aligned} \tag{3.51}$$

An alternative notation for this volume is the determinant

$$V = \det \underline{\underline{Q}} = \left| \underline{\underline{Q}} \right| \tag{3.52}$$

where the matrix $\underline{\underline{Q}}$ has the three basis vectors as its columns:

$$\underline{\underline{Q}} = \begin{bmatrix} A_x & B_x & C_x \\ A_y & B_y & C_y \\ A_z & B_z & C_z \end{bmatrix}. \tag{3.53}$$

Definition 34 (Higher-dimensional determinants)**.** Eq. 3.51 serves as the definition of a determinant for a 3×3 matrix. Higher-dimensional determinants (and volumes) can be specified via "expansion by minors":

$$\det \underline{\underline{A}} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij} \tag{3.54}$$

where a_{ij} is the matrix element of $\underline{\underline{A}}$ and M_{ij} is the minor of $\underline{\underline{A}}$; the determinant of the matrix formed by removing row i and column j from $\underline{\underline{A}}$.

Remark. This is an incredibly awkward formulation of the determinant to use.

Theorem 7. *The determinant of a product is the product of determinants:*

$$\det \underline{\underline{A}}\,\underline{\underline{B}} = \det \underline{\underline{A}} \det \underline{\underline{B}}. \tag{3.55}$$

Proof. See Arfken. □

Theorem 8. *The determinant of an inverse is the inverse of the determinant:*

$$\det \left(\underline{\underline{A}}^{-1} \right) = \frac{1}{\det \left(\underline{\underline{A}} \right)}. \tag{3.56}$$

Proof. A matrix inverse implies

$$\underline{\underline{A}}\,\underline{\underline{A}}^{-1} = \underline{\underline{I}}. \tag{3.57}$$

The determinant of this is

$$\det \left(\underline{\underline{A}}\,\underline{\underline{A}}^{-1} \right) = \det \left(\underline{\underline{I}} \right). \tag{3.58}$$

However, $\det \underline{\underline{I}} = 1$, therefore

$$\det \left(\underline{\underline{A}}^{-1} \right) = \frac{1}{\det \left(\underline{\underline{A}} \right)}. \tag{3.59}$$

□

Lemma 6. *The determinant of a similarity transform is equal to the determinant of the untransformed operator*

$$\det\left(\underline{\underline{B}}\,\underline{\underline{A}}\,\underline{\underline{B}}^{-1}\right) = \det\underline{\underline{A}}. \tag{3.60}$$

Proof. Follows from above. □

Theorem 9 (A useful definition of determinants)**.** *The determinant of an operator is the product of its eigenvalues.*

Proof. Let $\underline{\underline{W}}$ be the matrix of eigenvectors which diagonalizes the matrix $\underline{\underline{A}}$ (per Thm. 6) such that

$$\underline{\underline{W}}\,\underline{\underline{A}}\,\underline{\underline{W}}^{\dagger} = \underline{\underline{\Lambda}}. \tag{3.61}$$

Then, by Lem. 6, we have

$$\det\underline{\underline{A}} = \det\left(\underline{\underline{W}}\,\underline{\underline{A}}\,\underline{\underline{W}}^{\dagger}\right) = \det\underline{\underline{\Lambda}}. \tag{3.62}$$

However, the definition of the determinant (Eq. 3.54) tells us

$$\det\underline{\underline{\Lambda}} = \prod_i^n \lambda_i \tag{3.63}$$

for diagonal matrices such as $\underline{\underline{\Lambda}}$. □

Corollary 1. *An immediate consequence of this is* $\det\underline{\underline{A}} = 0$ *if* $\underline{\underline{A}}$ *is singular.*

Remark. Determinants are frequently found in geometry (see above), coordinate transforms (coming up), classical mechanics, and statistics (see below).

Example 16 (What is a partition function (or normalization in Bayesian methods)?)**.** One of the most perplexing aspects of statistical mechanics is getting an intuitive feel of partition functions (and therefore free energies). Determinants help us realize that this quantity is really just a weighted volume – much like the volume of a higher-dimensional geometric object.

Suppose we have a system of N harmonic oscillators (think back to normal modes) such that the thermally-scaled energy of the system is

$$H(\underline{x}) = \frac{1}{2}\underline{x}^{\dagger}\underline{\underline{A}}\,\underline{x}, \tag{3.64}$$

where $\underline{x}$ is the positions of the oscillators and $\underline{\underline{A}}$ is the $N \times N$ (SPD) stiffness matrix. The canonical partition function is proportional to the integral

$$Z = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\underline{x}^{\dagger}\underline{\underline{A}}\,\underline{x}\right)dx_1\cdots dx_N. \tag{3.65}$$

We use the Hubbard-Stratonovich transformation to relate this integral to a determinant. First, let $\underline{y} = \underline{\underline{W}}\,\underline{x}$, where $\underline{\underline{W}}$ is the orthonormal basis that diagonalizes $\underline{\underline{A}}$:

$$\underline{\underline{W}}^{\dagger}\underline{\underline{A}}\,\underline{\underline{W}} = \underline{\underline{\Lambda}}. \tag{3.66}$$

Then we can write

$$
\begin{aligned}
Z &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\underline{y}^{\dagger}\,\underline{\underline{W}}^{\dagger}\,\underline{\underline{A}}\,\underline{\underline{W}}\,\underline{y}\right) dy_1 \cdots dy_N \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\underline{y}^{\dagger}\,\underline{\underline{\Lambda}}\,\underline{y}\right) dy_1 \cdots dy_N \\
&= \prod_{i}^{N} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\lambda_i y_i^2\right) dy_i \qquad (3.67)
\end{aligned}
$$

However, this is just a product of Guassians, which can be written as

$$
Z = (2\pi)^{N/2} \prod_{i}^{N} \lambda_i^{-1/2}. \qquad (3.68)
$$

Using the above definition of the determinant (Thm. 9) we get

$$
Z = \sqrt{\frac{(2\pi)^N}{\det \underline{\underline{A}}}}. \qquad (3.69)
$$

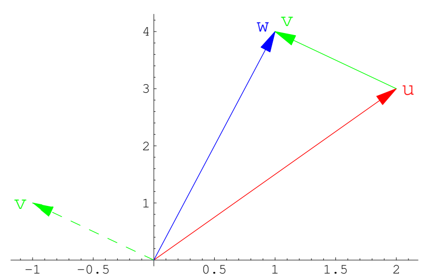
What is $\det \underline{\underline{A}}$ or $\det \underline{\underline{A}}^{-1}$?

4. Application to orthogonal coordinate transforms

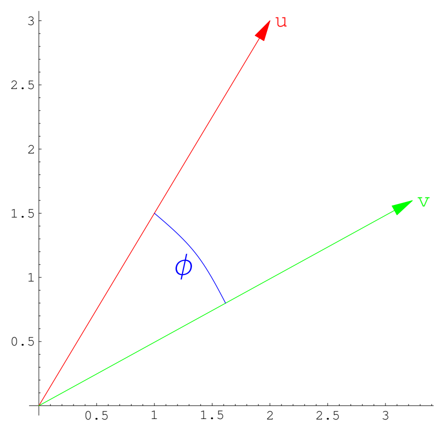
Remark. We'll consider three types of coordinate transformations: translation, rotation, and scaling as matrix-vector operations.

Definition 35 (Scaling)**.** Scaling is equivalent to scalar multiplication. Suppose we have a vector $\underline{v} = (x, y)$ with length $\|\underline{v}\| = \sqrt{x^2 + y^2}$. If we multiply by a scalar α we get a new vector $\underline{w} = (\alpha x, \alpha y)$ with length $\|\underline{w}\| = \alpha\sqrt{x^2 + y^2}$.

Definition 36 (Translation)**.** Translation is equivalent to vector addition:



Definition 37 (Rotation)**.** As we've seen before, rotation is equivalent to matrix multiplication:



Consider the vector $\underline{u} = (u_x, u_y)$ rotated by an angle ϕ to give the transformed vector $\underline{v} = (v_x, v_y)$. Basic trigonometry gives

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} u_x \\ u_y \end{bmatrix}. \tag{4.1}$$

Example 17. How do we "undo" a rotation represented by

$$\underline{\underline{A}} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}? \tag{4.2}$$

Solution. The matrix inverse!

$$\underline{\underline{A}}^{-1} = \underline{\underline{A}}^{\dagger} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}. \tag{4.3}$$

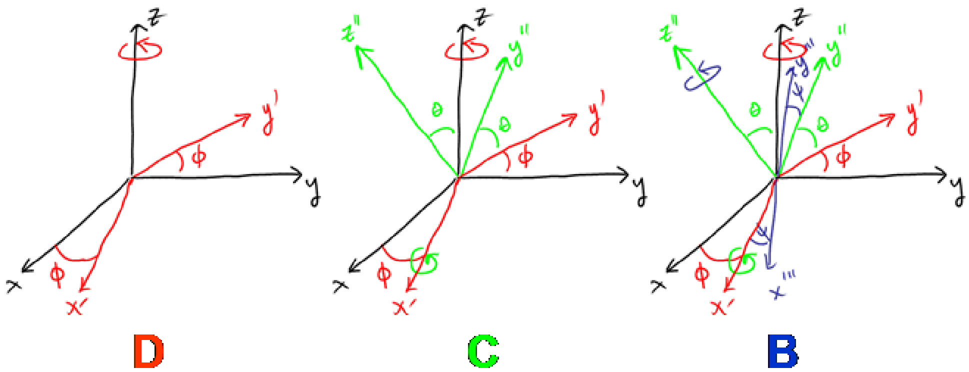
Check that this is the inverse:

$$\begin{aligned} \underline{\underline{A}}\,\underline{\underline{A}}^{-1} &= \begin{bmatrix} \cos^2\phi + \sin^2\phi & -\sin\phi\cos\phi + \sin\phi\cos\phi \\ \sin\phi\cos\phi - \sin\phi\cos\phi & \cos^2\phi + \sin^2\phi \end{bmatrix} \\ &= \underline{\underline{I}}. \end{aligned} \tag{4.4}$$

□

Remark. This is an *orthogonal transformation* because the adjoint of the transformation matrix is its inverse.

Definition 38 (Euler angles)**.** Euler angles represent 3-dimensional rotations in terms of a series of matrix operations:



In particular, a rotation is a series of three simple rotations

$$\underline{\underline{A}} = \underline{\underline{B}}\,\underline{\underline{C}}\,\underline{\underline{D}}. \tag{4.5}$$

The first rotation is around the z-axis:

$$\underline{\underline{D}} = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.6}$$

The second rotation is around the (new) x-axis:

$$\underline{\underline{C}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}. \tag{4.7}$$

The last rotation is around the (new) z-axis (again):

$$\underline{\underline{B}} = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.8}$$

Clearly, this transformation is invertible:

$$\underline{\underline{A}}^{-1} = \underline{\underline{D}}^{-1}\underline{\underline{C}}^{-1}\underline{\underline{B}}^{-1} = \underline{\underline{D}}^{\dagger}\underline{\underline{C}}^{\dagger}\underline{\underline{B}}^{\dagger} = \underline{\underline{A}}^{\dagger}. \tag{4.9}$$

In other words, this is also an orthogonal transformation.

Example 18. Suppose you have two identical proteins, one rotated with respect to the other. How might you find the rotation matrix to superimpose these proteins? What if they were slightly different?

Solution. Let 3 points of the first protein be $\underline{s}$, $\underline{t}$, $\underline{u}$ and the same points in the second protein be $\underline{s}'$, $\underline{t}'$, $\underline{u}'$. Then we have

$$\begin{aligned} \underline{s}' &= \underline{\underline{A}}\,\underline{s} \\ \underline{t}' &= \underline{\underline{A}}\,\underline{t} \\ \underline{v}' &= \underline{\underline{A}}\,\underline{v} \end{aligned}$$

or

$$\underline{\underline{X}}' = \underline{\underline{A}}\,\underline{\underline{X}} \tag{4.10}$$

where $\underline{\underline{X}}$ is the matrix with $\underline{s}$, $\underline{t}$, $\underline{u}$ as columns. We $\underline{\underline{X}}$ is invertible when the determinant is non-zero. What does this tell us about the volume or planarity of the points? Assuming $\det \underline{\underline{X}} \neq 0$, we have

$$\underline{\underline{A}} = \underline{\underline{X}}'\,\underline{\underline{X}}^{-1}. \tag{4.11}$$

If the proteins are slightly different, then we are better off using least squares:

$$\frac{\partial}{\partial \phi}\sum_i \left|\underline{x}'_i - \underline{\underline{A}}\,\underline{x}_i\right|^2 = 0 \tag{4.12}$$

$$\frac{\partial}{\partial \theta}\sum_i \left|\underline{x}'_i - \underline{\underline{A}}\,\underline{x}_i\right|^2 = 0 \tag{4.13}$$

$$\frac{\partial}{\partial \psi}\sum_i \left|\underline{x}'_i - \underline{\underline{A}}\,\underline{x}_i\right|^2 = 0. \tag{4.14}$$

□

5. Application to general coordinate transforms

Definition 39 (Jacobian matrices)**.** In general, we can write (linear) coordinate transforms as

$$x'_i = \sum_j a_{ij} x_j \tag{5.1}$$

$$x_i = \sum_j b_{ij} x'_j \tag{5.2}$$

which defines two matrices $\underline{\underline{A}}$ and $\underline{\underline{B}}$ where $\underline{\underline{A}} = \underline{\underline{B}}^{-1}$, if the inverses exist. These are called Jacobian matrices with elements:

$$a_{ij} = \frac{\partial x'_i}{\partial x_j} \tag{5.3}$$

$$b_{ij} = \frac{\partial x_i}{\partial x'_j}. \tag{5.4}$$

Definition 40 (Jacobian determinants)**.** Clearly, the invertibility of the coordinate transformation depends on det $\underline{\underline{A}}$ and det $\underline{\underline{B}}$ being non-zero. However, these determinants (known as Jacobian determinants) also give the ratio of unit volumes in the coordinate systems:

$$d\,\underline{x}' = \det\,\underline{\underline{A}}\,d\,\underline{x}\,. \tag{5.5}$$

Example 19. Suppose we decide to solve a problem in the inverse prolate spheroidal coordinate system, where:

$$x = \frac{\sinh\eta\sin\theta\cos\psi}{\cosh^2\eta - \sin^2\theta} \tag{5.6}$$

$$y = \frac{\sinh\eta\sin\theta\sin\psi}{\cosh^2\eta - \sin^2\theta} \tag{5.7}$$

$$z = \frac{\cosh\eta\cos\theta}{\cosh^2\eta - \sin^2\theta} \tag{5.8}$$

and $0 \leq \eta < \infty$, $0 \leq \theta \leq \pi$, $0 \leq \psi \leq 2\pi$. How would you write a volume integral of a function $f(\eta,\theta,\psi)$ in this coordinate system?

Solution. Build the matrix

$$\underline{\underline{J}} = \begin{bmatrix} \frac{\partial x}{\partial\eta} & \frac{\partial x}{\partial\theta} & \frac{\partial x}{\partial\psi} \\ \frac{\partial y}{\partial\eta} & \frac{\partial y}{\partial\theta} & \frac{\partial y}{\partial\psi} \\ \frac{\partial z}{\partial\eta} & \frac{\partial z}{\partial\theta} & \frac{\partial z}{\partial\psi} \end{bmatrix}. \tag{5.9}$$

and calculate its determinant

$$\det\,\underline{\underline{J}} = \frac{4\left(\cos 2\theta - \cosh 2\eta\right)\sin\theta\sinh\eta}{\left(\cos 2\theta + \cosh 2\eta\right)^3}. \tag{5.10}$$

The integral is then

$$\int_0^\infty \int_0^\pi \int_0^{2\pi} f(\eta,\theta,\psi)\frac{4\left(\cos 2\theta - \cosh 2\eta\right)\sin\theta\sinh\eta}{\left(\cos 2\theta + \cosh 2\eta\right)^3}d\eta d\theta d\psi. \tag{5.11}$$

□

Remark. Metric tensors offer a much easier route to developing surface integrals, volume integrals, and various derivatives in all sorts of coordinate systems.

Definition 41 (Metric tensor)**.** Let $\{q_i\}$ be a set of generalized coordinates. Then the distance to a point can be written as

$$ds^2 = \sum_i \sum_j g_{ij} q_i q_j \tag{5.12}$$

where g_{ij} is an element of the metric tensor $\underline{\underline{G}}$ and is describes how distances change with coordinate system:

$$g_{ij} = \sum_k \frac{\partial x_k}{\partial q_i}\frac{\partial x_k}{\partial q_j}. \tag{5.13}$$

Lemma 7. *For an orthogonal coordinate system the metric tensor is diagonal*

$$g_{ij} = h_i \delta_{ij} \tag{5.14}$$

with elements h_i called scale factors*.*

Proof. See Arfken. □

Definition 42 (Vector integrals)**.** With these definitions, we can write any volume integral for an orthogonal coordinate system as

$$\int_{\Omega} u(q_1, q_2, q_3) h_1 h_2 h_3 dq_1 dq_2 dq_3 \tag{5.15}$$

and any surface integral as

$$\int_{\partial\Omega} u(q_1, q_2) h_1 h_2 dq_1 dq_2. \tag{5.16}$$

Definition 43 (Vector derivatives)**.** Likewise, we can define general expressions for vector calculus derivative operations, including the gradient

$$(\nabla u)_i = \frac{\partial u}{\partial s_i} = \frac{1}{h_i}\frac{\partial u}{\partial q_i}, \tag{5.17}$$

the divergence

$$\nabla \cdot \underline{u} = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial\left(u_1 h_2 h_3\right)}{\partial q_1} + \frac{\partial\left(u_2 h_1 h_3\right)}{\partial q_2} + \frac{\partial\left(u_3 h_1 h_2\right)}{\partial q_3}\right], \tag{5.18}$$

the curl

$$\nabla \times \underline{u} = \det\left(\begin{bmatrix} h_1 & h_2 & h_3 \\ \frac{\partial}{\partial q_1} & \frac{\partial}{\partial q_2} & \frac{\partial}{\partial q_3} \\ h_1 u_1 & h_2 u_2 & h_3 u_3 \end{bmatrix}\right), \tag{5.19}$$

and the Laplacian:

$$\nabla^2 u = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial u}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_1 h_3}{h_2}\frac{\partial u}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial u}{\partial q_3}\right)\right]. \tag{5.20}$$

Remark. Although it seems a bit obtuse at first, this formulation makes dealing with strange (orthogonal) coordinate systems much easier.

Example 20 (Cartesian coordinates)**.** Our coordinates are

$$q_1 = x, \; -\infty < q_1 < \infty \tag{5.21}$$

$$q_2 = y, \; -\infty < q_2 < \infty \tag{5.22}$$

$$q_3 = z, \; -\infty < q_3 < \infty. \tag{5.23}$$

and our scale factors are

$$h_1 = 1 \tag{5.24}$$

$$h_2 = 1 \tag{5.25}$$

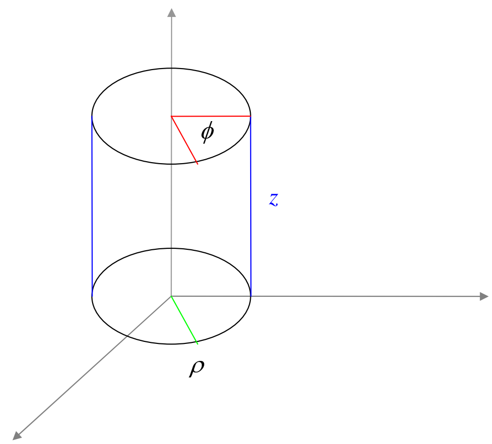
$$h_3 = 1. \tag{5.26}$$

The distance metric is

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{5.27}$$

and all the vector calculus operations follow as usual.

Example 21 (Circular cylindrical coordinates)**.** The circular cylindrical coordinate system has the form:



with coordinates

$$z = z, \quad -\infty < z < \infty \tag{5.28}$$
$$\rho = \sqrt{x^2 + y^2}, \quad 0 \le \rho < \infty \tag{5.29}$$
$$\phi = \operatorname{atan}\left(\frac{y}{x}\right) \quad 0 \le \phi \le 2\pi \tag{5.30}$$

or

$$x = \rho \cos \phi \tag{5.31}$$
$$y = \rho \sin \phi \tag{5.32}$$
$$z = z. \tag{5.33}$$

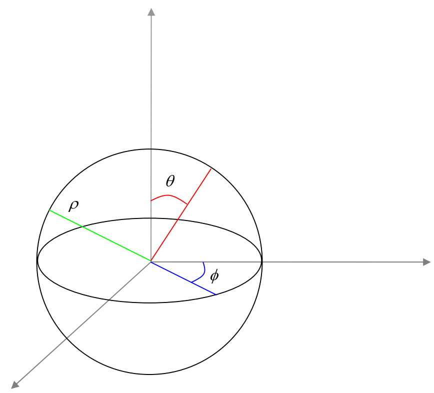
The scale factors are

$$h_\rho = 1 \tag{5.34}$$
$$h_\phi = \rho \tag{5.35}$$
$$h_z = 1. \tag{5.36}$$

The vector calculus operations follow from the equations above; as an example:

$$\begin{aligned} \nabla^2 u &= \frac{1}{\rho}\left[\frac{\partial}{\partial z}\left(\rho \frac{\partial u}{\partial z}\right) + \frac{\partial}{\partial \rho}\left(\rho \frac{\partial u}{\partial \rho}\right) + \frac{\partial}{\partial \phi}\left(\frac{1}{\rho}\frac{\partial u}{\partial \phi}\right)\right] \\ &= \frac{\partial^2 u}{\partial z} + \frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial u}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \phi^2}. \end{aligned} \tag{5.37}$$

Example 22 (Spherical polar coordinates)**.** The spherical polar coordinate system has the form:



with coordinates

$$r = \sqrt{x^2+y^2+z^2}, \quad 0 \le r < \infty \tag{5.38}$$
$$\theta = \text{acos}\left(\frac{z}{r}\right), \quad 0 \le \theta \le \pi \tag{5.39}$$
$$\phi = \text{atan}\left(\frac{y}{x}\right), \quad 0 \le \phi \le 2\pi \tag{5.40}$$

or

$$x = r\sin\theta\cos\theta \tag{5.41}$$
$$y = r\sin\theta\sin\theta \tag{5.42}$$
$$z = r\cos\theta. \tag{5.43}$$

The scale factors are

$$h_r = 1 \tag{5.44}$$
$$h_\theta = r \tag{5.45}$$
$$h_\phi = r\sin\theta. \tag{5.46}$$

As an example, this gives a Laplacian of

$$\begin{aligned}\nabla^2 u &= \frac{1}{r^2\sin\theta}\left[\frac{\partial}{\partial r}\left(r^2\sin\theta\frac{\partial u}{\partial r}\right)+\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right)+\frac{\partial}{\partial\phi}\left(\frac{1}{\sin\theta}\frac{\partial u}{\partial\phi}\right)\right]\\ &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\left(\frac{\partial^2 u}{\partial\phi^2}\right).\end{aligned} \tag{5.47}$$

6. Other applications

Remark. The rest of this course will draw upon basic linear algebra methods... I've intentionally left out linear systems of algebraic equations in this introduction because they don't lead to much insight into what basic linear operations *mean*. Don't forget: everything we've talked about works for (finite dimensional) spaces of functions!!!

7. Further reading

My favorite book for all things linear algebra is: Halmos PR, *Finite-Dimensional Vector Spaces*. An excellent reference for different coordinate systems is: Moon P, Spencer DE, *Field Theory Handbook*.