Math 115A - Week 8
Textbook sections: 5.2, 6.1
Topics covered:

- Characteristic polynomials
- Tests for diagonalizability
- Inner products
- Inner products and length

* * * * *

Characteristic polynomials

- Let A be an $n \times n$ matrix. Last week we introduced the *characteristic polynomial*
$$f(\lambda) = \det(A - \lambda I)$$
of that matrix; this is a polynomial in λ. For instance, if
$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$
then
$$\begin{aligned} f(\lambda) &= \det \begin{pmatrix} a-\lambda & b & c \\ d & e-\lambda & f \\ g & h & i-\lambda \end{pmatrix} \\ &= (a-\lambda)((e-\lambda)(i-\lambda)-fh)) - b(d(i-\lambda)-gf) + c(dh-(e-\lambda)g), \end{aligned}$$
which simplifies to some degree 3 polynomial in λ (we think of a, b, c, d, e, f, g as just constant scalars). Last week we saw that the zeroes of this polynomial give the eigenvalues of A.

- As you can see, the characteristic polynomial looks pretty messy. But in the special case of a diagonal matrix, e.g

$$A = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

the characteristic polynomial is quite simple, in fact

$$f(\lambda) = (a-\lambda)(b-\lambda)(c-\lambda)$$

(why?). This has zeroes when $\lambda = a, b, c$, and so the eigenvalues of this matrix are a, b, and c.

- **Lemma 1.** Let A and B be similar matrices. Then A and B have the same characteristic polynomial.

- An algebraist would phrase this as: "the characteristic polynomial is invariant under similarity".

- **Proof.** Since A and B are similar, we have $B = QAQ^{-1}$ for some invertible matrix Q. So the characteristic polynomial of B is

$$\begin{aligned} \det(B-\lambda I) &= \det(QAQ^{-1}-\lambda I) \\ &= \det(QAQ^{-1}-Q\lambda IQ^{-1}) \\ &= \det(Q(A-\lambda I)Q^{-1}) \\ &= \det(Q)\det(A-\lambda I)\det(Q^{-1}) \\ &= \det(Q)\det(Q^{-1})\det(A-\lambda I) \\ &= \det(QQ^{-1}(A-\lambda I)) \\ &= \det(A-\lambda I) \end{aligned}$$

and hence the characteristic polynomials are the same. $\square$

- Now let's try to understand the characteristic polynomial for general matrices. Let $P_1(\mathbf{R})$ be all the polynomials $a\lambda + b$ of degree at most 1; we shall make the free variable λ instead of x. Note that all the entries in the matrix $A - \lambda I$ lie in $P_1(\mathbf{R})$.

- **Lemma 2.** Let B be an $n \times n$ matrix, all of whose entries lie in $P_1(\mathbf{R})$. Then $\det(B)$ lies in $P_n(\mathbf{R})$ (i.e. $\det(B)$ is a polynomial in λ of degree at most n).

- **Proof.** We prove this by induction on n. When $n = 1$ the claim is trivial, since a 1×1 matrix with an entry in $P_1(\mathbf{R})$ looks like $B = (a\lambda + b)$, and clearly $\det(B) = a\lambda + b \in P_n(\mathbf{R})$.

  Now let's suppose inductively that $n > 1$, and that we have already proved the lemma for $n - 1$. We expand $\det(B)$ using cofactor expansion along some row or column (it doesn't really matter which row or column we use). This expands $\det(B)$ as an (alternating-sign) sum of expressions, each of which is the product of an entry of B, and a cofactor of B. The entry of B is in $P_1(\mathbf{R})$, while the cofactor of B is in $P_{n-1}(\mathbf{R})$ by the induction hypothesis. So each term in $\det(B)$ is in $P_n(\mathbf{R})$, and so $\det(B)$ is also in $P_n(\mathbf{R})$. This finishes the induction. $\square$

- From this lemma we see that $f(\lambda)$ lies in $P_n(\mathbf{R})$, i.e it is a polynomial of degree at most n. But we can be more precise. In fact the characteristic polynomial in general looks a lot like the characteristic polynomial of a diagonal matrix, except for an error which is a polynomial of degree at most $n - 2$:

- **Lemma 3.** Let $n \geq 2$. Let A be the $n \times n$ matrix

$$A := \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{pmatrix}.$$

  Then we have

$$f(\lambda) = (A_{11} - \lambda)(A_{22} - \lambda) \dots (A_{nn} - \lambda) + g(\lambda)$$

  where $g(\lambda) \in P_{n-2}(\mathbf{R})$.

- **Proof.** Again we induct on n. If $n = 2$ then $f(\lambda) = (A_{11} - \lambda)(A_{22} - \lambda) - A_{12}A_{21}$ (why?) and so the claim is true with $g := -A_{12}A_{21} \in P_0(\mathbf{R})$. Now suppose inductively that $n > 2$, and the claim has already been proven for $n - 1$. We write out $f(\lambda)$ as

$$f(\lambda) = \begin{pmatrix} A_{11} - \lambda & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} - \lambda & \dots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} - \lambda \end{pmatrix}.$$

- Now we do cofactor expansion along the first row. The first term in this expansion is

$$(A_{11} - \lambda) \det \begin{pmatrix} A_{22} - \lambda & \dots & A_{2n} \\ \vdots & \vdots & \vdots \\ A_{n2} & \dots & A_{nn} - \lambda \end{pmatrix}.$$

But this determinant is just the characteristic polynomial of an $n-1 \times n-1$ matrix, and so by the induction hypothesis we have

$$\det \begin{pmatrix} A_{22} - \lambda & \dots & A_{2n} \\ \vdots & \vdots & \vdots \\ A_{n2} & \dots & A_{nn} - \lambda \end{pmatrix} = (A_{22}-\lambda) \dots (A_{nn}-\lambda) + \text{ something in } P_{n-3}(\mathbf{R}).$$

Thus the first term in the cofactor expansion is

$$(A_{11} - \lambda)(A_{22} - \lambda) \dots (A_{nn} - \lambda) + \text{ something in } P_{n-2}(\mathbf{R}).$$

(Why did the $P_{n-3}(\mathbf{R})$ become a $P_{n-2}(\mathbf{R})$ when multiplying by $(A_{11} - \lambda)$?).

- Now let's look at the second term in the cofactor expansion; this is

$$-A_{12} \det \begin{pmatrix} A_{21} & \dots & A_{2n} \\ \vdots & \vdots & \vdots \\ A_{n1} & \dots & A_{nn} - \lambda \end{pmatrix}.$$

We do cofactor expansion again on the second row of this $n-1 \times n-1$ determinant. We can expand this determinant as an alternating-sign sum of terms, which look like A_{2i} times some $n-2 \times n-2$ determinant. By Lemma 2, this $n-2 \times n-2$ determinant lies in $P_{n-2}(\mathbf{R})$, while A_{2i} is a scalar. Thus all the terms in this determinant lie in $P_{n-2}(\mathbf{R})$, and so the determinant itself must lie in $P_{n-2}(\mathbf{R})$ (recall that $P_{n-2}(\mathbf{R})$ is closed under addition and scalar multiplication). Thus this second term in the cofactor expansion lies in $P_{n-2}(\mathbf{R})$.

- A similar argument shows that the third, fourth, etc. terms in the cofactor expansion of $\det(A - \lambda I)$ all lie in $P_{n-2}(\mathbf{R})$. Adding up all these terms we obtain

$$\det(A - \lambda I) = (A_{11} - \lambda)(A_{22} - \lambda) \dots (A_{nn} - \lambda) + \text{ something in } P_{n-2}(\mathbf{R})$$

as desired. $\square$

- If we multiply out

$$(A_{11} - \lambda)(A_{22} - \lambda) \dots (A_{nn} - \lambda)$$

we get

$$(-\lambda)^n + (-\lambda)^{n-1}(A_{11} + A_{22} + \dots + A_{nn}) + \text{ stuff of degree at most } n-2$$

(why?). Note that $(A_{11} + \dots + A_{nn})$ is just the trace $\mathrm{tr}(A)$ of A. Thus from Lemma 3 we have

$$f(\lambda) = (-1)^n \lambda^n + (-1)^{n-1}\mathrm{tr}(A)\lambda^{n-1} + a_{n-2}\lambda^{n-2} + a_{n-3}\lambda^{n-3} + \dots + a_1\lambda + a_0$$

for some scalars $a_{n-2}, \dots, a_0$. These coefficients $a_{n-2}, \dots, a_0$ are quite interesting, but hard to compute. However, a_0 can be obtained by a simple trick: if we evaluate the above expression at 0, we get

$$f(0) = a_0,$$

but $f(0) = \det(A - 0I) = \det(A)$. We have thus proved the following result.

- **Theorem 4.** The characteristic polynomial $f(\lambda)$ of an $n \times n$ matrix A has the form

$$f(\lambda) = (-1)^n \lambda^n + (-1)^{n-1}\mathrm{tr}(A)\lambda^{n-1} + a_{n-2}\lambda^{n-2} + a_{n-3}\lambda^{n-3} + \dots + a_1\lambda + \det(A).$$

- Thus the characteristic polynomial encodes the trace and the determinant, as well as some additional information which we will not study further in this course.

- **Example.** The characteristic polynomial of the 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is

$$(a - \lambda)(d - \lambda) - bc = \lambda^2 - (a + d)\lambda + (ad - bc)$$

(why?). Note that $a + d$ is the trace and $ad - bc$ is the determinant.

- **Example.** The characteristic polynomial of the 3×3 matrix

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

is

$$(a-\lambda)(b-\lambda)(c-\lambda) = -\lambda^3 + (a+b+c)\lambda^2 - (ab+bc+ca)\lambda + abc.$$

Note that $a+b+c$ is the trace and abc is the determinant.

- Since the characteristic polynomial is of degree n and has a leading coefficient of -1, it is possible that it factors into n linear factors, i.e.

$$f(\lambda) = -(\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n)$$

for some scalars $\lambda_1, \ldots, \lambda_n$ in the field of scalars (which we will call F for a change... this F may be either $\mathbf{R}$ or $\mathbf{C}$). These scalars do not necessarily have to be distinct (i.e. we can have releated roots). If this is the case we say that f *splits over* F, or more simply that f *splits*.

- **Example.** The characteristic polynomial of the 2×2 matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is $\lambda^2 - 1$ (why?), which splits over the reals as $(\lambda-1)(\lambda-(-1))$. It also splits over the complex numbers because $+1$ and -1 are real numbers, and hence also complex numbers. On the other hand, the characteristic polynomial of

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

is $\lambda^2 + 1$, which doesn't split over the reals, but does split over the complexes as $(\lambda - i)(\lambda + i)$. Finally, the characteristic polynomial of

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

is λ^2, which splits over both the reals and the complexes as $(\lambda-0)(\lambda-0)$.

- **Example** The characteristic polynomial of a diagonal matrix will always split. For instance the characteristic polynomial of

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

is

$$(a-\lambda)(b-\lambda)(c-\lambda) = -(\lambda-a)(\lambda-b)(\lambda-c).$$

- From the previous example, and Lemma 1, we see that the characteristic polynomial of any diagonalizable matrix will always split (since diagonalizable matrices are similar to diagonal matrices). In particular, if the characteristic polynomial of a matrix *doesn't* split, then it can't be diagonalizable.

- **Example.** The matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

from an earlier example cannot be diagonalizable over the reals, because its characteristic polynomial does not split over the reals. (However, it can be diagonalized over the complex numbers; we leave this as an exercise).

- It turns out that the complex numbers have a significant advantage over the reals, in that polynomials *always* split:

- **Fundamental Theorem of Algebra.** Every polynomial splits over the complex numbers.

- This theorem is a basic reason why the complex numbers are so useful; unfortunately, the proof of this theorem is far beyond the scope of this course. (You can see a proof in Math 132, however).

* * * * *

Tests for diagonalizability

- Recall that an $n \times n$ matrix A is *diagonalizable* if there is an invertible matrix Q and a diagonal matrix D such that $A = QDQ^{-1}$. It is often useful to know when a matrix can be diagonalized. We already know one such characterization: A is diagonalizable if and only if there is a basis of $\mathbf{R}^n$ which consists entirely of eigenvectors of A. Equivalently:

- **Lemma 5.** An $n \times n$ matrix A is diagonalizable if and only if one can find n linearly independent vectors $v_1, v_2, \ldots, v_n$ in $\mathbf{R}^n$, such that each vector v_j is an eigenvector of A.

- This is because n linearly independent vectors in $\mathbf{R}^n$ automatically form a basis of $\mathbf{R}^n$.

- It is thus important to know when the eigenvectors of A are linearly independent. Here is one useful test:

- **Proposition 6.** Let A be an $n \times n$ matrix. Let $v_1, v_2, \ldots, v_k$ be eigenvectors of A with eigenvalues $\lambda_1, \ldots, \lambda_k$ respectively. Suppose that the eigenvalues $\lambda_1, \ldots, \lambda_k$ are all *distinct*. Then the vectors $v_1, \ldots, v_k$ are linearly independent.

- **Proof.** Suppose for contradiction that $v_1, \ldots, v_k$ were not independent, i.e. there was some scalars $a_1, \ldots, a_k$, not all equal to zero, such that

$$a_1 v_1 + a_2 v_2 + \ldots + a_k v_k = 0.$$

At least one of the a_j is non-zero; without loss of generality we may assume that a_1 is non-zero.

- Now we use a trick to eliminate v_k: We apply $(A - \lambda_k I)$ to both sides of this equation. Using the fact that $A - \lambda_k I$ is linear, we obtain

$$a_1(A - \lambda_k I)v_1 + a_2(A - \lambda_k I)v_2 + \ldots + a_k(A - \lambda_k I)v_k = 0.$$

But observe that

$$(A - \lambda_k I)v_1 = Av_1 - \lambda_k v_1 = \lambda_1 v_1 - \lambda_k v_1 = (\lambda_1 - \lambda_k)v_1$$

and more generally

$$(A - \lambda_k I)v_j = (\lambda_j - \lambda_k)v_j.$$

In particular we have

$$(A - \lambda_k I)v_k = 0.$$

Putting this all together, we obtain

$$a_1(\lambda_1 - \lambda_k)v_1 + a_2(\lambda_2 - \lambda_k)v_2 + \ldots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = 0.$$

Now we eliminate v_{k-1} by applying $A - \lambda_{k-1}I$ to both sides of the equation. Arguing as before, we obtain

$$a_1(\lambda_1 - \lambda_k)(\lambda_1 - \lambda_{k-1})v_1 + a_2(\lambda_2 - \lambda_k)(\lambda_2 - \lambda_{k-1})v_2 + \ldots$$

$$+a_{k-2}(\lambda_{k-2} - \lambda_k)(\lambda_{k-2} - \lambda_{k-1})v_{k-2} = 0.$$

We then eliminate v_{k-2}, then v_{k-3}, and so forth all the way down to eliminating v_2, until we obtain

$$a_1(\lambda_1 - \lambda_k)(\lambda_1 - \lambda_{k-1}) \ldots (\lambda_1 - \lambda_2)v_1 = 0.$$

But since the λ_i are all distinct, and a_1 is non-zero, this forces v_1 to equal zero. But this contradicts the definition of eigenvector (eigenvectors are not allowed to be zero). Thus the vectors $v_1, \ldots, v_k$ must have been linearly independent. □

- Proposition 5 holds for linear transformations as well as matrices: see Theorem 5.10 of the textbook.

- **Corollary 6** Let A be an $n \times n$ matrix. If the characteristic polynomial of A splits into n *distinct* factors, then A is diagonalizable.

- **Proof.** By assumption, the characteristic polynomial $f(\lambda)$ splits as

$$f(\lambda) = -(\lambda - \lambda_1) \ldots (\lambda - \lambda_n)$$

for some distinct scalars $\lambda_1, \ldots, \lambda_n$. Thus we have n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. For each eigenvalue λ_j let v_j be an eigenvector with that eigenvalue, then by Proposition 5 $v_1, \ldots, v_n$ are linearly independent, and hence by Lemma 4 A is diagonalizable. □

- **Example** Consider the matrix

$$A = \begin{pmatrix} 1 & -2 \\ 1 & 4 \end{pmatrix}.$$

The characteristic polynomial here is

$$f(\lambda) = (1-\lambda)(4-\lambda) + 2 = \lambda^2 - 5\lambda + 6 = (\lambda-2)(\lambda-3),$$

so the characteristic polynomial splits into n distinct factors (regardless of whether our scalar field is the reals or the complexes). So we know that A is diagonalizable. (If we actually wanted the explicit diagonalization, we would find the eigenvalues (which are 2,3) and then some eigenvectors, and use the previous week's notes).

- To summarize what we know so far: if the characteristic polynomial doesn't split, then we can't diagonalize the matrix; while if it does split into distinct factors, then we can diagonalize the matrix. There is still a remaining case in which the characteristic function splits, but into repeated factors. Unfortunately this case is much more complicated; the matrix may or may not be diagonalizable. For instance, the matrix

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

has a characteristic polynomial of $(\lambda - 2)^2$ (why?), so it splits but not into distinct linear factors. It is clearly diagonalizable (indeed, it is diagonal). On the other hand, the matrix

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$$

has the same characteristic polynomial of $(\lambda - 2)^2$ (why?), but it turns out not to be diagonalizable, for the following reason. If it were diagonalizable, then we could find a basis of $\mathbf{R}^n$ which consists entirely of eigenvectors. But since the only root of the characteristic polynomial is 2, the only eigenvalue is 2. Now let's work out what the eigenvectors are. Since the only eigenvalue is 2, we only need to look in the eigenspace with eigenvalue 2. We have to solve the equation

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2 \begin{pmatrix} x \\ y \end{pmatrix},$$

i.e. we have to solve the system of equations

$$2x + y = 2x; \quad 2y = 2y.$$

The general solution of this system occurs when $y = 0$ and x is arbitrary, so the eigenspace with eigenvalue 2 is just the x-axis. But the vectors from this eigenspace are not enough to span all of $\mathbf{R}^2$, so we cannot find a basis of eigenvectors. Thus this matrix is not diagonalizable.

- The moral of this story is that, while the characteristic polynomial does carry a large amount of information, it does not completely solve the problem of whether a matrix is diagonalizable or not. However, even when the characteristic polynomial is inconclusive, it is still possible to determine whether a matrix is diagonalizable or not by computing its eigenspaces and seeing if it is possible to make a basis consisting entirely of eigenvectors. We will not pursue the full solution of the diagonalization problem here, but defer it to 115B (where you will learn about two more tools to study diagonalization - the minimal polynomial and the Jordan normal form).

- **One last example.** Consider the matrix

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix};$$

this is the same matrix as the previous example but we attach another row and column, and add a 3. (This is not a diagonal matrix, but is an example of a *block-diagonal matrix*: see this week's homework for more information). The characteristic polynomial here is

$$f(\lambda) = -(\lambda - 2)^2(\lambda - 3)$$

(why?), so the eigenvalues are 2 and 3. To find the eigenspace with eigenvalue 2, we solve the equation

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2 \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

and a little bit of work shows that the general solution to this equation occurs when $y = z = 0$ and x is arbitrary, thus the eigenspace is just the x-axis $\{ \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} : x \in \mathbf{R} \}$. Similarly the eigenspace with eigenvalue 3 is the z-axis. But this is not enough eigenvectors to span $\mathbf{R}^3$ (the 2-eigenspace only contributes one linearly independent eigenvector, and the 3-eigenspace contributes only one linearly independent eigenvector, whereas we need three linearly independent eigenvectors in order to span $\mathbf{R}^3$.

* * * * *

Inner product spaces

- We now leave matrices and eigenvalues and eigenvectors for the time being, and begin a very different topic - the concept of an inner product space.

- Up until now, we have been preoccupied with *vector spaces* and various things that we can do with these vector spaces. If you recall, a vector space comes equipped with only two basic operations: addition and scalar multiplication. These operations have already allowed us to introduce many more concepts (bases, linear transformations, etc.) but they cannot do everything that one would like to do in applications.

- For instance, how does one compute the length of a vector? In $\mathbf{R}^2$ or $\mathbf{R}^3$ one can use Pythagoras's theorem to work out the length, but what about, say, a vector in $P_3(\mathbf{R})$? What is the length of $x^3 + 3x^2 + 6$? It turns out that such spaces do not have an inherent notion of length: you can add and scalar multiply two polynomials, but we have not given any rule to determine the length of a polynomial. Thus, vector spaces are not equipped to handle certain geometric notions such as length (or angle, or orthogonality, etc.)

- To resolve this, mathematicians have introduced several "upgraded" versions of vector spaces, in which you can not only add and scalar multiply vectors, but can also compute lengths, angles, inner products, etc. One particularly common such "upgraded" vector space is something called an *inner product space*, which we will now discuss. (There are also *normed vector spaces*, which have a notion of length but not angle; *topological vector spaces*, which have a notion of convergence but not length; and if you wander into infinite dimensions then there are slightly fancier things such as *Hilbert spaces* and *Banach spaces*. Then there are *vector algebras*, where you can multiply vectors with other vectors to get more vectors. Then there are hybrids of these notions, such as *Banach algebras*, which are a certain type of infinite-dimensional vector algebra. None of these will be covered in this course; they are mostly graduate level topics).

- The problem with length is that it is not particularly linear: the length of a vector $v+w$ is not just the length of v plus the length of w. However, in $\mathbf{R}^2$ or $\mathbf{R}^3$ we can rewrite the length of a vector v as the square root of the *dot product* $v \cdot v$. Unlike length, the dot product is linear in the sense that $(v + v') \cdot w = v \cdot w + v' \cdot w$ and $v \cdot (w + w') = v \cdot w + v \cdot w'$, with a similar rule for scalar multiplication. (Actually, to be precise, the dot product is considered *bilinear* rather than linear, just as the determinant is considered multilinear, because it has two inputs v and w, instead of just one for linear transformations).

- Thus, the idea behind an inner product space is to introduce length indirectly, by means of something called an *inner product*, which is a generalization of the dot product. Depending on whether the field of scalars is real or complex, we have either a real inner product space or a complex inner product space. Complex inner product spaces are similar to real ones, except the complex conjugate operation $z \mapsto \overline{z}$ makes an appearance. Here's a clue why: the length $|z|$ of a complex number $z = a + bi$, is not the square root of $z \cdot z$, but is instead the square root of $z \cdot \overline{z}$.

- We will now use both real and complex vector spaces, and will try to take care to distinguish between the two. When we just say "vector space" without the modifier "real" or "complex", then the field of scalars might be either the reals or the complex numbers.

- **Definition** An *inner product space* is a vector space V equipped with an additional operation, called an *inner product*, which takes two vectors

$v, w \in V$ as input and returns a scalar $\langle v, w \rangle$ as output, which obeys the following three properties:

- (Linearity in the first variable) For any vectors $v, v', w \in V$ and any scalar c, we have $\langle v + v', w \rangle = \langle v, w \rangle + \langle v', w \rangle$ and $\langle cv, w \rangle = c\langle v, w \rangle$.

- (Conjugate symmetry) If v and w are vectors in V, then $\langle w, v \rangle$ is the complex conjugate of $\langle v, w \rangle$: $\langle w, v \rangle = \overline{\langle v, w \rangle}$.

- (Positivity) If v is a non-zero vector in V, then $\langle v, v \rangle$ is a positive real number: $\langle v, v \rangle > 0$.

- If the field of scalars is real, then every number is its own conjugate (e.g. $\overline{3} = 3$) and so the conjugate-symmetry property simplifies to just the symmetry property $\langle w, v \rangle = \langle v, w \rangle$.

- We now give some examples of inner product spaces.

- **$\mathbf{R}^n$ as an inner product space.** We already know that $\mathbf{R}^n$ is a real vector space. If we now equip $\mathbf{R}^n$ with the inner product equal to the *dot product*
$$\langle x, y \rangle := x \cdot y$$
i.e.
$$\langle (x_1, x_2, \ldots, x_n), (y_1, \ldots, y_n) \rangle = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_{j=1}^n x_j y_j$$
then we obtain an inner product space. For instance, we now have $\langle (1, 2), (3, 4) \rangle = 11$.

- To verify that we have an inner product space, we have to verify the linearity property, conjugate symmetry property, and the positivity property. To verify the linearity property, observe that
$$\langle x + x', y \rangle = (x + x') \cdot y = x \cdot y + x' \cdot y = \langle x, y \rangle + \langle x', y \rangle$$
and
$$\langle cx, y \rangle = (cx) \cdot y = c(x \cdot y) = c\langle x, y \rangle$$

while the conjugate symmetry follows since

$$\langle y, x\rangle = y \cdot x = x \cdot y = \langle x, y\rangle = \overline{\langle x, y\rangle}$$

(since the conjugate of a real number is itself. To verify the positivity property, observe that

$$\langle (x_1, x_2, \ldots, x_n), (x_1, \ldots, x_n)\rangle = x_1^2 + x_2^2 + \ldots + x_n^2$$

which is clearly positive if the $x_1, \ldots, x_n$ are not all zero.

- The difference between the dot product and the inner product is that the dot product is specific to $\mathbf{R}^n$, while the inner product is a more general concept and is applied to many other vector spaces.

- One can interpret the dot product $x \cdot y$ as measuring the amount of "correlation" or "interaction" between x and y; the longer that x and y are, and the more that they point in the same direction, the larger the dot product becomes. If x and y point in opposing directions then the dot product is negative, while if x and y point at right angles then the dot product is zero. Thus the dot product combines both the length of the vectors, and their angle (as can be seen by the famous formula $x \cdot y = |x||y|\cos\theta$ but easier to work with than either length or angle because it is (bi-)linear (while length and angle individually are not linear quantities).

- **$\mathbf{R}^n$ as an inner product space II.** One doesn't have to use the dot product as the inner product; other dot products are possible. For instance, one could endow $\mathbf{R}^n$ with the non-standard inner product

$$\langle x, y\rangle' := 10x \cdot y,$$

so for instance $\langle (1,2), (3,4)\rangle' = 110$. While this is not the standard inner product, it still obeys the three properties of linearity, conjugate symmetry, and positivity (why?), so this is still an inner product (though to avoid confusion we have labeled it as $\langle , \rangle'$ instead of $\langle , \rangle$. The situation here is similar to bases of vector spaces; a vector space such as $\mathbf{R}^n$ can have a standard basis but also have several non-standard bases (for instance, we could multiply every vector in the standard basis by 10), and the same is often true of inner products. However in the vast majority of cases we will use a standard inner product.

- More generally, we can multiply any inner product by a positive constant, and still have an inner product.

- **R as an inner product space.** A special case of the previous example of $\mathbf{R}^n$ with the standard inner product occurs when $n = 1$. Then our inner product space is just the real numbers, and the inner product is given by the ordinary product: $\langle x, y \rangle := xy$. For instance $\langle 2, 3 \rangle = 6$. Thus, plain old multiplication is itself an example of an inner product space.

- **C as an inner product space.** Now let's look at the complex numbers $\mathbf{C}$, which is a one-dimensional *complex* vector space (so the field of scalars is now $\mathbf{C}$). Here, we could reason by analogy with the previous example and guess that $\langle z, w \rangle := zw$ would be an inner product, but this does not obey either the conjugate-symmetry property or the positivity property: if z were a complex number, then $\langle z, z \rangle = z^2$ would not necessarily be a positive real number (or even a real number); for instance $\langle i, i \rangle = -1$.

- To fix this, the correct way to define an inner product on $\mathbf{C}$ is to set $\langle z, w \rangle := z\overline{w}$; in other words we have to conjugate the second factor. This inner product is now linear in the first variable (why?) and conjugate-symmetric (why?). To verify positivity, observe that $\langle a + bi, a + bi \rangle = (a + bi)(a - bi) = a^2 + b^2$ which will be a positive real number if $a + bi$ is non-zero.

- $\mathbf{C}^n$ **as an inner product space.** Now let's look at the complex vector space
$$\mathbf{C}^n := \{(z_1, z_2, \ldots, z_n) : z_1, z_2, \ldots, z_n \in \mathbf{C}\}.$$
This is just like $\mathbf{R}^n$ but with the scalars being complex instead of real; for instance $(3, 1 + i, 3i)$ would lie in $\mathbf{C}^3$ but it wouldn't be a vector in $\mathbf{R}^3$. We can define an inner product here by
$$\langle (z_1, z_2, \ldots, z_n), (w_1, w_2, \ldots, w_n) \rangle := z_1\overline{w_1} + z_2\overline{w_2} + \ldots + z_n\overline{w_n};$$
note how this definition is a hybrid of the $\mathbf{R}^n$ inner product and the $\mathbf{C}$ inner product. This is an inner product space (why?).

- **Functions as an inner product space.** Consider $C([0,1];\mathbf{R})$, the space of continuous real-valued functions from the interval $[0,1]$ to $\mathbf{R}$. This is an infinite-dimensional real vector space, containing such functions as $\sin(x)$, x^2+3, $1/(x+1)$, and so forth. We can define an inner product on this space by defining

$$\langle f, g\rangle := \int_0^1 f(x)g(x)\ dx;$$

for instance,

$$\langle x+1, x^2\rangle = \int_0^1 (x+1)x^2\ dx = (\frac{x^4}{4} + \frac{x^3}{3})|_0^1 = \frac{1}{4} + \frac{1}{3} = \frac{7}{12}.$$

Note that we need the continuity property in order to make sure that this integral actually makes sense (as opposed to diverging to infinity or otherwise doing something peculiar). One can verify fairly easily that this is an inner product space; we just give parts of this verification. One of the things we have to show is that

$$\langle f_1 + f_2, g\rangle = \langle f_1, g\rangle + \langle f_2, g\rangle,$$

but this follows since the left-hand side is

$$\int_0^1 (f_1(x)+f_2(x))g(x)\ dx = \int_0^1 f_1(x)g(x)\ dx + \int_0^1 f_2(x)g(x)\ dx = \langle f_1, g\rangle + \langle f_2, g\rangle.$$

To verify positivity, observe that

$$\langle f, f\rangle = \int_0^1 f(x)^2\ dx.$$

The function $f(x)^2$ is always non-negative, and if f is not the zero function on $[0,1]$, then $f(x)^2$ must be strictly positive for some $x \in [0,1]$. Thus there is a strictly positive area under the graph of $f(x)^2$, and so $\int_0^1 f(x)^2\ dx > 0$.

- One can view this example as an infinite-dimensional version of the finite-dimensional inner product space example of $\mathbf{R}^n$. To see this, let N be a very large number. Remembering that integrals can be approximated by Riemann sums, we have (non-rigorously) that

$$\int_0^1 f(x)g(x)\ dx \approx \sum_{j=1}^{N} f(\frac{j}{N})g(\frac{j}{N})\frac{1}{N},$$

or in other words

$$\langle f, g\rangle \approx \frac{1}{N}(f(\frac{1}{N}), f(\frac{2}{N}), \ldots, f(\frac{N}{N})) \cdot (g(\frac{1}{N}), g(\frac{2}{N}), \ldots, g(\frac{N}{N})),$$

and the right-hand side resembles an example of the inner product on $\mathbf{R}^N$ (admittedly there is an additional factor of $\frac{1}{N}$, but as observed before, putting a constant factor in the definition of an inner product just gives you another inner product.

- **Functions as an inner product space II.** Consider $C([-1,1];\mathbf{R})$, the space of continuous real-valued functions on $[-1,1]$. Here we can define an inner product as

$$\langle f, g\rangle := \int_{-1}^1 f(x)g(x)\ dx.$$

Thus for instance

$$\langle x+1, x^2\rangle = \int_{-1}^1 (x+1)x^2\ dx = (\frac{x^4}{4} + \frac{x^3}{3})|_{-1}^1 = \frac{2}{3}.$$

Note that this inner product of $x+1$ and x^2 was different from the inner product of $x+1$ and x^2 given in the previous example! Thus it is important, when dealing with functions, to know exactly what the domain of the functions is, and when dealing with inner products, to know exactly which inner product one is using - confusing one inner product for another can lead to the wrong answer! To avoid confusion, one sometimes labels the inner product with some appropriate subscript, for instance the inner product here might be labeled $\langle, \rangle_{C([-1,1];\mathbf{R})}$ and the previous one labeled $\langle, \rangle_{C([0,1];\mathbf{R})}$.

- **Functions as an inner product space III.** Now consider $C([0,1];\mathbf{C})$, the space of continuous complex-valued functions from the interval $[0,1]$

to $\mathbf{C}$; this includes such functions as $\sin(x)$, $x^2 + ix - 3 + i$, $i/(x - i)$, and so forth. Note that while the range of this function is complex, the domain is still real, so x is still a real number. This is a infinite-dimensional complex vector space (why?). We can define an inner product on this space as

$$\langle f, g \rangle := \int_0^1 f(x)\overline{g(x)}\ dx.$$

Thus, for instance

$$\langle x^2, x + i \rangle = \int_0^1 x^2(x - i)\ dx = (\frac{x^4}{4} - \frac{ix^3}{3})|_0^1 = \frac{1}{4} - \frac{i}{3}.$$

This can be easily verified to be an inner product space. For the positivity, observe that

$$\langle f, f \rangle = \int_0^1 f(x)\overline{f(x)}\ dx = \int_0^1 |f(x)|^2\ dx.$$

Even though $f(x)$ can be any complex number, $|f(x)|^2$ must be a non-negative real number, and an argument similar to that for real functions shows that $\int_0^1 |f(x)|^2\ dx$ is a positive real number when f is not the zero function on $[0, 1]$.

- **Polynomials as an inner product space.** The inner products in the above three examples work on functions. Since polynomials are a special instance of functions, the above inner products are also inner products on polynomials. Thus for instance we can give $P_3(\mathbf{R})$ the inner product

  $$\langle f, g \rangle := \int_0^1 f(x)g(x)\ dx,$$

  so that for instance

  $$\langle x, x^2 \rangle = \int_0^1 x^3\ dx = \frac{x^4}{4}|_0^1 = \frac{1}{4}.$$

  Or we could instead give $P_3(\mathbf{R})$ a different inner product

  $$\langle f, g \rangle := \int_{-1}^1 f(x)g(x)\ dx,$$

so that for instance

$$\langle x, x^2 \rangle = \int_{-1}^{1} x^3 \, dx = \frac{x^4}{4}|_{-1}^{1} = 0.$$

Unfortunately, we have here a large variety of inner products and it is not obvious what the "standard" inner product should be. Thus whenever we deal with polynomials as an inner product space we shall be careful to specify exactly which inner product we will use. However, we can draw one lesson from this example, which is that if V is an inner product space, then any subspace W of V is also an inner product space. (Note that if the properties of linearity, conjugate symmetry, and positivity hold for the larger space V, then they will automatically hold for the smaller space W (why?)).

- **Matrices as an inner product space.** Let $M_{m \times n}(\mathbf{R})$ be the space of real matrices with m rows and n columns; a typical element is

$$A = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \ldots & A_{mn} \end{pmatrix}.$$

- This is a mn-dimensional real vector space. We can turn this into an inner product space by defining the inner product

$$\langle A, B \rangle := \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij};$$

i.e. for every row and column we multiply the corresponding entries of A and B together, and then sum. For instance,

$$\langle \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \rangle = 1 \times 5 + 2 \times 6 + 3 \times 7 + 4 \times 8 = 70.$$

It is easy to verify that this is also an inner product; note how similar this is to the standard $\mathbf{R}^n$ inner product. This inner product can also be written using transposes and traces:

$$\langle A, B \rangle = \mathrm{tr}(AB^t).$$

To see this, note that AB^t is an $m \times m$ matrix, with the top left entry being $A_{11}B_{11} + A_{12}B_{12} + \ldots + A_{1n}B_{1n}$, the second entry on the diagonal being $A_{21}B_{21} + A_{22}B_{22} + \ldots + A_{2n}B_{2n}$, and so forth down the diagonal (why?). Adding these together we obtain $\langle A, B\rangle$. For instance,

$$\mathrm{tr}(\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}^t) = \mathrm{tr} \begin{pmatrix} 1 \times 5 + 2 \times 6 & 1 \times 7 + 2 \times 8 \\ 3 \times 5 + 4 \times 5 & 3 \times 7 + 4 \times 8 \end{pmatrix}$$

$$= 1 \times 5 + 2 \times 6 + 3 \times 7 + 4 \times 8 = \langle \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}\rangle.$$

- **Matrices as an inner product space II.** Let $M_{m\times n}(\mathbf{C})$ be the space of complex matrices with m rows and n columns; this is an mn-dimensional complex vector space. This is an inner product space with inner product
$$\langle A, B\rangle := \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}\overline{B_{ij}}.$$

- It is not hard to verify that this is indeed an inner product space. Unlike the previous example, the inner product is not given by the formula $\langle A, B\rangle = \mathrm{tr}(AB^t)$, however there is a very similar formula. Define the *adjoint* $B^\dagger$ of a matrix to be the complex conjugate of the transpose B^t; i.e. $B^\dagger$ is the same matrix as B^t but with every entry replaced by its complex conjugate. For instance,
$$\begin{pmatrix} 1+2i & 3+4i \\ 5+6i & 7+8i \\ 9+10i & 11+12i \end{pmatrix}^\dagger = \begin{pmatrix} 1-2i & 5-6i & 9-10i \\ 3-4i & 7-8i & 11-12i \end{pmatrix}.$$
The adjoint is the complex version of the transpose; it is completely unrelated to the adjugate matrix in the previous week's notes. It is easy to verify that $\langle A, B\rangle = \mathrm{tr}(AB^\dagger)$.

- (Optional remarks) To summarize: many of the vector spaces we have encountered before, can be upgraded to inner product spaces. As we shall see, the additional capabilities of inner product spaces can be useful in many applications when the more basic capabilities of vector spaces are not enough. On the other hand, inner products add complexity (and possible confusion, if there is more than one choice of inner product available) to a problem, and in many situations (for instance, in the last six weeks of material) they are unnecessary. So it is sometimes better to just deal with bare vector spaces, with no inner product attached; it depends on the situation. (A more subtle reason why sometimes adding extra structure can be bad, is because it reduces the amount of symmetry in a situation; it is relatively easy for a transformation to be *linear* (i.e. it preserves the vector space structures of addition and scalar multiplication) but it is much harder to be *isometric* (which means that it not only preserves addition and scalar multiplication, but also inner products as well.). So if one insists on dealing with inner products all the time, then one loses a lot of symmetries, because there is more structure to preserve, and this can sometimes make a problem appear harder than it actually is. Some of the deepest advances in physics, for instance, particularly in relativity and quantum mechanics, were only possible because the physicists removed a lot of unnecessary structure from their models (e.g. in relativity they removed separate structures for space and time, keeping only something called the spacetime metric), and then gained so much additional symmetry that they could then use those symmetries to discover new laws of physics (e.g. Einstein's law of gravitation).)

- Some basic properties of inner products:

- From the linearity and conjugate symmetry properties it is easy to see that $\langle v, w + w' \rangle = \langle v, w \rangle + \langle v, w' \rangle$ and $\langle v, cw \rangle = \overline{c} \langle v, w \rangle$ for all vectors v, w, w' and scalars c (why?) Note that when you pull a scalar c out of the second factor, it gets conjugated, so be careful about that. (Another way of saying this is that the inner product is *conjugate linear*, rather than linear, in the second variable. Because the inner product is linear in the first variable and only sort-of-linear in the second, it is sometimes said that the inner product is *sesquilinear* (sesqui is Latin for "one and a half").

- The inner product of 0 with anything is 0: $\langle 0, v \rangle = \langle v, 0 \rangle = 0$. (This is an easy consequence of the linearity (or conjugate linearity) - Why?). In particular, $\langle 0, 0 \rangle = 0$. Thus, by the positivity property, $\langle v, v \rangle$ is positive if v is non-zero and zero if v is zero. In particular, if we ever know that $\langle v, v \rangle = 0$, we can deduce that v itself is 0.

Inner products and length

- Once you have an inner product space, you can then define a notion of length:

- **Definition** Let V be an inner product space, and let v be a vector of V. Then the *length* of v, denoted $\|v\|$, is given by the formula $\|v\| := \sqrt{\langle v, v \rangle}$. (In particular, $\langle v, v \rangle = \|v\|^2$).

- **Example** In $\mathbf{R}^2$ with the standard inner product, the vector $(3, 4)$ has length
$$\|(3,4)\| = \sqrt{\langle (3,4), (3,4) \rangle} = \sqrt{3^2 + 4^2} = 5.$$

- If instead we use the non-standard inner product $\langle x, y \rangle = 10x \cdot y$, then the length is now
$$\|(3,4)\| = \sqrt{\langle (3,4), (3,4) \rangle} = \sqrt{10(3^2 + 4^2)} = 5\sqrt{10}.$$
Thus the notion of length depends very much on what inner product you choose (although in most cases this will not be an issue since we will use the standard inner product).

- From the positivity property we see that every non-zero vector has a positive length, while the zero vector has zero length. Thus $\|v\| = 0$ if and only if $v = 0$.

- If c is a scalar, then
$$\|cv\| = \sqrt{\langle cv, cv \rangle} = \sqrt{c\bar{c}\langle v, v \rangle} = \sqrt{|c|^2 \|v\|^2} = |c|\|v\|.$$

- **Example** In a complex vector space, the vector $(3 + 4i)v$ is five times as long as v. The vector $-v$ has exactly the same length as v.

- The inner product in some special cases can be expressed in terms of length. We already know that $\langle v, v \rangle = \|v\|^2$. More generally, if w is a positive scalar multiple of v (so that v and w are parallel and in the same direction), then $w = cv$ for some positive real number c, and

$$\langle v, w\rangle = \langle v, cv\rangle = c\langle v, v\rangle = c\|v\|^2 = \|v\|\|cv\| = \|v\|\|w\|,$$

i.e. when v and w are pointing in the same direction then the inner product is just the product of the norms. In the other extreme, if w is a negative scalar multiple of v, then $w = -cv$ for some positive c, and

$$\langle v, w\rangle = \langle v, -cv\rangle = -c\langle v, v\rangle = -c\|v\|^2$$

$$= -\|v\|| - c|\|v\| = -\|v\|\| - cv\| = -\|v\|\|w\|,$$

and so the inner product is negative the product of the norms. In general the inner product lies in between these two extremes:

- **Cauchy-Schwarz inequality** Let V be an inner product space. For any $v, w \in V$, we have

$$|\langle v, w\rangle| \leq \|v\|\|w\|.$$

- **Proof.** If $w = 0$ then both sides are zero, so we can assume that $w \neq 0$. From the positivity property we know that $\langle v, v\rangle \geq 0$. More generally, for any scalars a, b we know that $\langle av + bw, av + bw\rangle \geq 0$. But

$$\begin{aligned} \langle av + bw, av + bw\rangle \quad &= a\langle v, av + bw\rangle + b\langle w, av + bw\rangle \\ &= a\overline{a}\langle v, v\rangle + a\overline{b}\langle v, w\rangle + b\overline{a}\langle w, v\rangle + b\overline{b}\langle w, w\rangle \\ &= |a|^2\|v\|^2 + a\overline{b}\langle v, w\rangle + b\overline{a}\overline{\langle v, w\rangle} + |b|^2\|w\|^2. \end{aligned}$$

Since $\langle av + bw, av + bw\rangle \geq 0$ for any choice of scalars a, b, we thus have

$$|a|^2\|v\|^2 + a\overline{b}\langle v, w\rangle + b\overline{a}\overline{\langle v, w\rangle} + |b|^2\|w\|^2 \geq 0$$

for any choice of scalars a, b. We now select a and b in order to obtain some cancellation. Specifically, we set

$$a := \|w\|^2; b := -\langle v, w\rangle.$$

Then we see that

$$\|w\|^4\|v\|^2 - \|w\|^2\overline{\langle v, w\rangle}\langle v, w\rangle - \langle v, w\rangle\|w\|^2\overline{\langle v, w\rangle} + |\langle v, w\rangle|^2\|w\|_2^2 \geq 0;$$

this simplifies to

$$\|w\|^4\|v\|^2 \geq \|w\|^2|\langle v, w\rangle|^2.$$

Dividing by $\|w\|^2$ (recall that w is non-zero, so that $\|w\|$ is non-zero) and taking square roots we obtain the desired inequality. $\square$

- Thus for real vector spaces, the inner product $\langle v, w\rangle$ always lies somewhere between $+\|v\|\|w\|$ and $-\|v\|\|w\|$. For complex vector spaces, the inner product $\langle v, w\rangle$ can lie anywhere in the disk centered at the origin with radius $\|v\|\|w\|$. For instance, in the complex vector space $\mathbf{C}$, if $v = 3 + 4i$ and $w = 4 - 3i$ then $\langle v, w\rangle = v\overline{w} = 25i$, while $\|v\| = 5$ and $\|w\| = 5$.

- The Cauchy-Schwarz inequality is extremely useful, especially in analysis; it tells us that if one of the vectors v, w have small length then their inner product will also be small (unless of course the other vector has very large length).

- Another fundamental inequality concerns the relationship between length and vector addition. It is clear that length is not linear: the length of $v + w$ is not just the sum of the length of v and the length of w. For instance, in $\mathbf{R}^2$, if $v := (1, 0)$ and $w = (0, 1)$ then $\|v + w\| = \|(1, 1)\| = \sqrt{2} \neq 1 + 1 = \|v\| + \|w\|$. However, we do have

- **Triangle inequality** Let V be an inner product space. For any $v, w \in V$, we have
$$\|v + w\| \leq \|v\| + \|w\|.$$

- **Proof.** To prove this inequality, we can square both sides (note that this is OK since both sides are non-negative):
$$\|v + w\|^2 \leq (\|v\| + \|w\|)^2.$$
The left-hand side is
$$\langle v + w, v + w\rangle = \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle.$$
The quantities $\langle v, v\rangle$ and $\langle w, w\rangle$ are just $\|v\|^2$ and $\|w\|^2$ respectively. From the Cauchy-Schwarz inequality, the two quantities $\langle v, w\rangle$ and $\langle w, v\rangle$ have absolute value at most $\|v\|\|w\|$. Thus
$$\langle v + w, v + w\rangle \leq \|v\|^2 + \|v\|\|w\| + \|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2$$

as desired. $\square$

- The reason this is called the triangle inequality is because it has a natural geometric interpretation: if one calls two sides of a triangle v and w, so that the third side is $v + w$, then the triangle inequality says that the length of the third side is less than or equal to the sum of the lengths of the other two sides. In other words, a straight line has the shortest distance between two points (at least when compared to triangular alternatives).

- The triangle inequality has a couple of variants. Here are a few:

$$\|v - w\| \leq \|v\| + \|w\|$$

$$\|v + w\| \geq \|v\| - \|w\|$$

$$\|v + w\| \geq \|w\| - \|v\|$$

$$\|v - w\| \geq \|v\| - \|w\|$$

$$\|v - w\| \geq \|w\| - \|v\|$$

  Thus for instance, if v has length 10, and w has length 3, then both $v+w$ and $v-w$ have length somewhere between 7 and 13. (Can you see this geometrically?). These inequalities can be proven in a similar manner to the original triangle inequality, or alternatively one can start with the original triangle inequality and do some substitutions (e.g. replace w by $-w$, or replace v by $v - w$, on both sides of the inequality; try this!).