(6.2) Angles and Orthogonality

Recall that we define the angle between two vectors in R^3 by $\cos\theta = \dfrac{\mathbf{u}\cdot\mathbf{v}}{||\mathbf{u}||\,||\mathbf{v}||}$. We want to make the same definition for vectors in inner product spaces, so we would define $\cos\theta = \dfrac{\langle \mathbf{u},\mathbf{v}\rangle}{||\mathbf{u}||\,||\mathbf{v}||}$. But there is a potential problem: what if for some inner product it turned out that $\langle \mathbf{u},\mathbf{v}\rangle > ||\mathbf{u}||\,||\mathbf{v}||$? Then $\cos\theta > 1$, but that's not possible. However, we can prove that in an inner product space, this cannot happen.

Cauchy-Schwarz inequality. For all vectors $\mathbf{u}$ and $\mathbf{v}$, we have

$$|\langle \mathbf{u},\mathbf{v}\rangle| \le ||\mathbf{u}||\,||\mathbf{v}||$$

The proof of this is clever and interesting. It relies on the fact that $||\mathbf{u}||^2 = \langle \mathbf{u},\mathbf{u}\rangle$ for any vector $\mathbf{u}$. Now $||t\mathbf{u}+\mathbf{v}||^2 \ge 0$ for any real number t. (The norm of anything is nonnegative, but we've squared it for good measure.) This is

$$\begin{aligned} ||t\mathbf{u}+\mathbf{v}||^2 &= \langle t\mathbf{u}+\mathbf{v}, t\mathbf{u}+\mathbf{v}\rangle \\ &= t^2\langle \mathbf{u},\mathbf{u}\rangle + t\langle \mathbf{u},\mathbf{v}\rangle + t\langle \mathbf{v},\mathbf{u}\rangle + \langle \mathbf{v},\mathbf{v}\rangle \\ &= t^2||\mathbf{u}||^2 + 2t\langle \mathbf{u},\mathbf{v}\rangle + ||\mathbf{v}||^2 \end{aligned}$$

Now this is supposed to be nonnegative. If we write $a = ||\mathbf{u}||^2$, $b = 2\langle \mathbf{u},\mathbf{v}\rangle$ and $c = ||\mathbf{v}||^2$ we have

$$at^2 + bt + c \ge 0$$

for all t. This means either the quadratic equation $at^2+bt+c=0$ has no real solutions or exactly one solution. The quadratic formula is

$$t = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$$

so it can be seen that there are no real solutions exactly when the discriminant b^2-4ac is negative, and exactly one real solution exactly when it is zero. So we know $b^2-4ac \le 0$. This is

$$(2\langle \mathbf{u},\mathbf{v}\rangle)^2 - 4\,||\mathbf{u}||^2\,||\mathbf{v}||^2 \le 0$$

or

$$(2\langle \mathbf{u},\mathbf{v}\rangle)^2 \le 4\,||\mathbf{u}||^2\,||\mathbf{v}||^2$$

If we take square roots and divide by 2, we get $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}|| \, ||\mathbf{v}||$. (If $x^2 \leq c^2$ then $|x| \leq c$, assuming $c \geq 0$.)

This verifies the Cauchy-Schwarz inequality in R^n, by the way, something that we didn't prove back in Chapter 4.

With the Cauchy-Schwarz inequality in hand, it is now easy to prove the Triangle Inequality. All we do is:

$$\begin{aligned} ||\mathbf{u}+\mathbf{v}||^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= ||\mathbf{u}||^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + ||\mathbf{v}||^2 \\ &\leq ||\mathbf{u}||^2 + 2\,||\mathbf{u}||\,||\mathbf{v}|| + ||\mathbf{v}||^2 \qquad \text{(C-S ineq.)} \\ &= (||\mathbf{u}|| + ||\mathbf{v}||)^2 \end{aligned}$$

so by taking square roots, we have

$$||\mathbf{u}+\mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||\,.$$

(If $a^2 \leq b^2$ and a and b are both nonnegative, then $a < b$.)

The reason why we need the Triangle Inequality is that it makes our definition of distance, $d(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$, behave correctly. If you go from point A to point B and then to point C, the distance you cover should be at least as large as if you go from point A directly to point C.

We are now equipped to talk about angles and distances in settings where you might not think such geometrical quantities could make any sense.

Example. We can find the angle between two functions in a function space. Say $V = C[0, 2\pi]$. Recall this is the set of all continuous functions on the interval $[0, 2\pi]$. Let's try $f(x) = \cos 2x$ and $g(x) = \sin 3x$. We compute:

$$\begin{aligned} \langle f, g \rangle &= \int_0^{2\pi} \cos 2x \, \sin 3x \, dx = 0 \\ ||f||^2 = \langle f, f \rangle &= \int_0^{2\pi} \big(\cos 2x\big)^2 \, dx = \pi \\ ||g||^2 = \langle g, g \rangle &= \int_0^{2\pi} \big(\sin 3x\big)^2 \, dx = \pi \end{aligned}$$

So

$$\cos\theta = \frac{\langle f, g\rangle}{||f||\,||g||} = \frac{0}{\sqrt{\pi}\cdot\sqrt{\pi}} = 0$$

Of course, we didn't need to compute the norms of f and g once we discovered that $\langle f, g\rangle = 0$. But we have found $\cos\theta = 0$ so $\theta = 90°$. So in some mysterious sense, these functions are perpendicular to each other.

We now make a very basic and important definition: in an inner product space V, we say two vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$. This means $\cos\theta = 0$ where θ is the angle between the vectors, so it means $\theta = 90°$. That is, the vectors are perpendicular. So in the last example, we had two functions that are orthogonal.

Orthogonal complements. Suppose W is a subspace of an inner product space V. We define $W^\perp$, the **orthogonal complement** of V, to be the set of all vectors in V that are orthogonal to W, that is, orthogonal to every vector in W. It is easy to verify that $W^\perp$ is also a subspace of V. For example, if $\mathbf{x}$ and $\mathbf{y}$ belong to $W^\perp$, then so does $\mathbf{x}+\mathbf{y}$. This is because $\langle \mathbf{x}+\mathbf{y}, \mathbf{w}\rangle = \langle \mathbf{x}, \mathbf{w}\rangle + \langle \mathbf{y}, \mathbf{w}\rangle = 0 + 0 = 0$ for every $\mathbf{w}$ in W.

We can visualize $W^\perp$ as follows. Say $V = R^3$ and W is a plane through the origin. Then $W^\perp$ ends up being the line through the origin that is perpendicular to the plane. If instead W is a line through the origin, then $W^\perp$ turns out to be the plane through the origin that is perpendicular to the line. (Recall all subspaces of R^3 are either lines through the origin, planes through the origin, or else $\{\mathbf{0}\}$, or all of R^3.)

Very often, a subspace of a vector space is given as the span of several vectors. So we might say, let W be the subspace of V that is spanned by the three vectors $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$. Or we might say, let W be the subspace of V that has basis $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$. We would like to be able to find $W^\perp$ in this circumstance. It turns out to be easy to do that. The trick is to form a matrix A whose rows are given by the vectors that span W. Then $W^\perp$ turns out to be exactly the same thing as the nullspace of A. This works because if $\mathbf{x}$ is in the nullspace of A, so it solves $A\mathbf{x} = \mathbf{0}$, then the dot product of each row of A with $\mathbf{x}$ is equal to 0. Rather than give a full-fledged proof of this, one simple example should be convincing. Look at a 3×3 matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Call each row of A a vector, so let $\mathbf{a}_1 = (a_{11}, a_{12}, a_{13})$, $\mathbf{a}_2 = (a_{21}, a_{22}, a_{23})$ and $\mathbf{a}_3 = (a_{31}, a_{32}, a_{33})$. Now suppose $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ is in the nullspace of A, which means $A\mathbf{x} = \mathbf{0}$. This is

$$\begin{aligned} A\mathbf{x} &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{x} \\ \mathbf{a}_2 \cdot \mathbf{x} \\ \mathbf{a}_3 \cdot \mathbf{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

so $\mathbf{a}_1 \cdot \mathbf{x} = 0$, $\mathbf{a}_2 \cdot \mathbf{x} = 0$ and $\mathbf{a}_3 \cdot \mathbf{x} = 0$.

Example. Suppose W is the subspace of R^5 spanned by the four vectors $(1, 3, -1, 8, 4)$, $(1, -1, -1, 4, 4)$, $(2, 0, 1, 1, 7)$ and $(0, 1, 1, -2, -1)$. We will find a basis for W and we will find a basis for $W^{\perp}$. Let A be the matrix whose rows are these vectors:

$$\begin{bmatrix} 1 & 3 & -1 & 8 & 4 \\ 1 & -1 & -1 & 4 & 4 \\ 2 & 0 & 1 & 1 & 7 \\ 0 & 1 & 1 & -2 & -1 \end{bmatrix}$$

We solve $A\mathbf{x} = \mathbf{0}$, which is a system that in augmented matrix form is given by

$$\left[\begin{array}{rrrrr|r} 1 & 3 & -1 & 8 & 4 & 0 \\ 1 & -1 & -1 & 4 & 4 & 0 \\ 2 & 0 & 1 & 1 & 7 & 0 \\ 0 & 1 & 1 & -2 & -1 & 0 \end{array}\right]$$

We put this into reduced row-echelon form:

$$\left[\begin{array}{rrrrr|r} 1 & 0 & 0 & 2 & 3 & 0 \\ 0 & 1 & 0 & 1 & -2 & 0 \\ 0 & 0 & 1 & -3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

This gives us a basis for W (and also for the row space of A): $(1, 0, 0, 2, 3)$, $(0, 1, 0, 1, -2)$, $(0, 0, 1, -3, 1)$. But if we write the system represented by the reduced matrix, we can find the nullspace of A. We have

$$\begin{array}{ccccccccc} x_1 & & & + & 2x_4 & + & 3x_5 & = & 0 \\ & x_2 & & + & x_4 & - & 2x_5 & = & 0 \\ & & x_3 & - & 3x_4 & + & x_5 & = & 0 \end{array}$$

We do the usual thing: solve for the lead variables and then set the free variables equal to certain parameters. We have

$$\begin{aligned} x_1 &= -2t - 3s \\ x_2 &= -t + 2s \\ x_3 &= 3t - s \\ x_4 &= t \\ x_5 &= s \end{aligned}$$

Now we write

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -2t - 3s \\ -t + 2s \\ 3t - s \\ t \\ s \end{bmatrix} = t \begin{bmatrix} -2 \\ -1 \\ 3 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} -3 \\ 2 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

which finally provides us with a basis for the nullspace of A, and therefore $W^\perp$: $(-2, -1, 3, 1, 0)$ and $(-3, 2, -1, 0, 1)$.

We may verify directly that these two vectors are orthogonal to each of the original four vectors that span W. For example, we can verify that the third vector in the set that spanned W is orthogonal to the second vector in the basis for $W^\perp$:

$$\langle (2, 0, 1, 1, 7), (-3, 2, -1, 0, 1) \rangle = 2(-3) + 0(2) + 1(-1) + 1(0) + 7(1) = 0.$$

(This inner product is just the ordinary dot product or inner product on R^5.)

By the way, the dimension of W turned out to be 3 (the basis for W has three vectors), and the dimension of $W^\perp$ turned out to be 2 (it has a basis of two vectors). Both of these spaces are subspaces of R^5. Notice $3 + 2 = 5$. This always happens: if W is a subspace of R^n then $\dim(W) + \dim(W^\perp) = n$. This is just the same fact we learned earlier, that rank + nullity $= n$ for an $m \times n$ matrix A.