LINEAR ALGEBRA. LECTURE 6

Orthogonal vectors and subspaces

In this lecture we learn what it means for vectors, bases and subspaces to be *orthogonal*. The symbol for this is $\perp$.

The "big picture" of this course is that the row space of a matrix' is orthogonal to its nullspace, and its column space is orthogonal to its left nullspace.

row space	column space
dimension r	dimension r
$\perp$	$\perp$
nullspace	left nullspace $N(A^T)$
dimension $n - r$	dimension $m - r$

Orthogonal vectors

Orthogonal is just another word for *perpendicular*. Two vectors are *orthogonal* if the angle between them is 90 degrees. If two vectors are orthogonal, they form a right triangle whose hypotenuse is the sum of the vectors. Thus, we can use the Pythagorean theorem to prove that *the dot product* $\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x}$ *is zero* exactly when $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. (The length squared $||\mathbf{x}||^2$ equals $\mathbf{x}^T\mathbf{x}$.)

Note that all vectors are orthogonal to the zero vector.

Orthogonal subspaces

Subspace S is *orthogonal* to subspace T means: every vector in S is orthogonal to every vector in T. The blackboard is not orthogonal to the floor; two vectors in the line where the blackboard meets the floor aren't orthogonal to each other.

In the plane, the space containing only the zero vector and any line through the origin are orthogonal subspaces. A line through the origin and the whole plane are never orthogonal subspaces. Two lines through the origin are orthogonal subspaces if they meet at right angles.

Nullspace is perpendicular to row space

The row space of a matrix is orthogonal to the nullspace, because $A\mathbf{x} = \mathbf{0}$ means the dot product of $\mathbf{x}$ with each row of A is 0. But then the product of $\mathbf{x}$ with any combination of rows of A must be 0.

The column space is orthogonal to the left nullspace of A because the row space of A^T is perpendicular to the nullspace of A^T.

In some sense, the row space and the nullspace of a matrix subdivide $\mathbb{R}^n$ into two perpendicular subspaces. For $A = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \end{bmatrix}$, the row space has dimension 1 and basis $\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$ and the nullspace has dimension 2 and is the plane through the origin perpendicular to the vector $\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$.

Not only is the nullspace orthogonal to the row space, their dimensions add up to the dimension of the whole space. We say that the nullspace and the row space are *orthogonal complements* in $\mathbb{R}^n$. The nullspace contains all the vectors that are perpendicular to the row space, and vice versa.

We could say that this is part two of the fundamental theorem of linear algebra. Part one gives the dimensions of the four subspaces, part two says those subspaces come in orthogonal pairs, and part three will be about orthogonal bases for these subspaces.

$N(A^TA) = N(A)$

Due to measurement error, $A\mathbf{x} = \mathbf{b}$ is often unsolvable if $m > n$. Our next challenge is to find the best possible solution in this case. The matrix A^TA plays a key role in this effort: the central equation is $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$.

We know that A^TA is square ($n \times n$) and symmetric. When is it invertible?

Suppose $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix}$. Then:

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & 30 \end{bmatrix}$$

is invertible. A^TA is not always invertible. In fact:

$$\begin{aligned} N(A^TA) &= N(A) \\ \text{rank of } A^TA &= \text{rank of } A. \end{aligned}$$

We conclude that A^TA is invertible exactly when A has independent columns.

Projections onto subspaces

Projections

If we have a vector $\mathbf{b}$ and a line determined by a vector $\mathbf{a}$, how do we find the point on the line that is closest to $\mathbf{b}$?

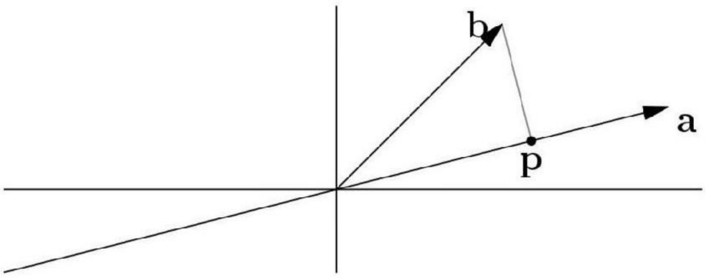

Figure 1: The point closest to $\mathbf{b}$ on the line determined by $\mathbf{a}$.

We can see from Figure 1 that this closest point $\mathbf{p}$ is at the intersection formed by a line through $\mathbf{b}$ that is orthogonal to $\mathbf{a}$. If we think of $\mathbf{p}$ as an approximation of $\mathbf{b}$, then the length of $\mathbf{e} = \mathbf{b} - \mathbf{p}$ is the error in that approximation.

We could try to find $\mathbf{p}$ using trigonometry or calculus, but it's easier to use linear algebra. Since $\mathbf{p}$ lies on the line through $\mathbf{a}$, we know $\mathbf{p} = x\mathbf{a}$ for some number x. We also know that $\mathbf{a}$ is perpendicular to $\mathbf{e} = \mathbf{b} - \mathbf{xa}$:

$$\begin{aligned} \mathbf{a}^T(\mathbf{b} - x\mathbf{a}) &= 0 \\ x\mathbf{a}^T\mathbf{a} &= \mathbf{a}^T\mathbf{b} \\ x &= \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}, \end{aligned}$$

and $\mathbf{p} = \mathbf{a}x = \mathbf{a}\dfrac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}$. Doubling $\mathbf{b}$ doubles $\mathbf{p}$. Doubling $\mathbf{a}$ does not affect $\mathbf{p}$.

Projection matrix

We'd like to write this projection in terms of a *projection matrix* P: $\mathbf{p} = P\mathbf{b}$.

$$\mathbf{p} = \mathbf{xa} = \frac{\mathbf{a}\mathbf{a}^T\mathbf{a}}{\mathbf{a}^T\mathbf{a}},$$

so the matrix is:

$$P = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}.$$

Note that $\mathbf{aa}^T$ is a three by three matrix, not a number; matrix multiplication is not commutative.

The column space of P is spanned by $\mathbf{a}$ because for any $\mathbf{b}$, $P\mathbf{b}$ lies on the line determined by $\mathbf{a}$. The rank of P is 1. P is symmetric. $P^2\mathbf{b} = P\mathbf{b}$ because the projection of a vector already on the line through $\mathbf{a}$ is just that vector. In general, projection matrices have the properties:

$$P^T = P \qquad \text{and} \qquad P^2 = P.$$

Why project?

As we know, the equation $A\mathbf{x} = \mathbf{b}$ may have no solution. The vector $A\mathbf{x}$ is always in the column space of A, and $\mathbf{b}$ is unlikely to be in the column space. So, we project $\mathbf{b}$ onto a vector $\mathbf{p}$ in the column space of A and solve $A\hat{\mathbf{x}} = \mathbf{p}$.

Projection in higher dimensions

In $\mathbb{R}^3$, how do we project a vector $\mathbf{b}$ onto the closest point $\mathbf{p}$ in a plane?

If $\mathbf{a}_1$ and $\mathbf{a}_2$ form a basis for the plane, then that plane is the column space of the matrix $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$.

We know that $\mathbf{p} = \hat{x}_1\mathbf{a}_1 + \hat{x}_2\mathbf{a}_2 = A\hat{\mathbf{x}}$. We want to find $\hat{\mathbf{x}}$. There are many ways to show that $\mathbf{e} = \mathbf{b} - \mathbf{p} = \mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the plane we're projecting onto, after which we can use the fact that $\mathbf{e}$ is perpendicular to $\mathbf{a}_1$ and $\mathbf{a}_2$:

$$\mathbf{a}_1^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \quad \text{and} \quad \mathbf{a}_2^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0.$$

In matrix form, $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$. When we were projecting onto a line, A only had one column and so this equation looked like: $a^T(\mathbf{b} - x\mathbf{a}) = \mathbf{0}$.

Note that $\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$ is in the nullspace of A^T and so is in the left nullspace of A. We know that everything in the left nullspace of A is perpendicular to the column space of A, so this is another confirmation that our calculations are correct.

We can rewrite the equation $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$ as:

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

When projecting onto a line, $A^T A$ was just a number; now it is a square matrix. So instead of dividing by $\mathbf{a}^T\mathbf{a}$ we now have to multiply by $(A^T A)^{-1}$

In n dimensions,

$$\begin{aligned} \hat{\mathbf{x}} &= (A^T A)^{-1} A^T\mathbf{b} \\ \mathbf{p} = A\hat{\mathbf{x}} &= A(A^T A)^{-1} A^T\mathbf{b} \\ P &= A(A^T A)^{-1} A^T. \end{aligned}$$

It's tempting to try to simplify these expressions, but if A isn't a square matrix we can't say that $(A^T A)^{-1} = A^{-1}(A^T)^{-1}$. If A does happen to be a square, invertible matrix then its column space is the whole space and contains $\mathbf{b}$. In this case P is the identity, as we find when we simplify. It is still true that:

$$P^T = P \qquad \text{and} \qquad P^2 = P.$$

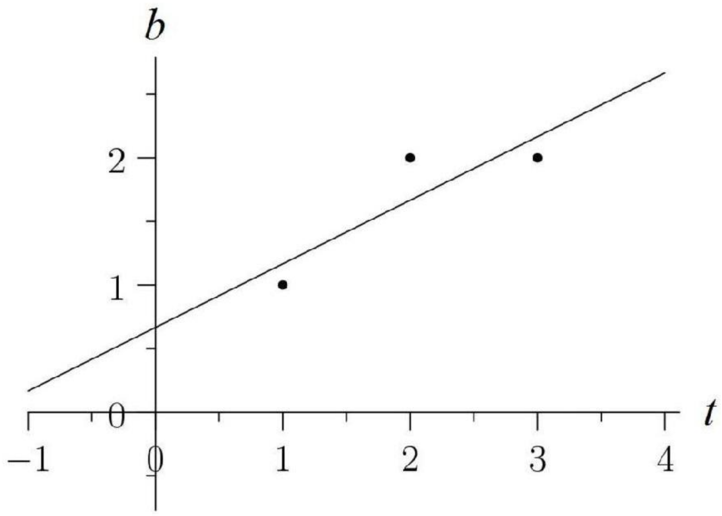

Figure 2: Three points and a line close to them.

Least Squares

Suppose we're given a collection of data points (t, b):

$$\{(1,1),(2,2),(3,2)\}$$

and we want to find the closest line $b = C + Dt$ to that collection. If the line went through all three points, we'd have:

$$\begin{aligned} C + D &= 1 \\ C + 2D &= 2 \\ C + 3D &= 2, \end{aligned}$$

which is equivalent to:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}}_{\mathbf{b}}.$$

In our example the line does not go through all three points, so this equation is not solvable. Instead we'll solve:

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$