Recall from Week 11

1. An inner product $\langle \cdot, \cdot \rangle : V \to \mathbb{R}$ on a real vector space V satisfies
   - $\langle u, v \rangle = \langle v, u \rangle$ for all $u, v \in V$.
   - Bilinearity
   - $\langle v, v \rangle \geq 0$ for all $v \in V$ (and $= 0$ only if $v = 0$).
2. Example: The ordinary scalar product on $\mathbb{R}^n$
3. Length: $||v|| = \sqrt{\langle v, v \rangle}$
   Distance: $d(u, v) = ||u - v||$
   Orthogonality: $u \perp v \iff \langle u, v \rangle = 0$.
4. Projection: for a pair of vectors u and v (with $u \neq 0$),

$$\text{proj}_u\, v = \frac{\langle u, v \rangle}{\langle u, u \rangle} u.$$

   Then $v - \text{proj}_u\, v$ is orthogonal to u, and $\text{proj}_u\, v$ is the nearest scalar multiple of u to v.

Orthogonal Bases (the Gram-Schmidt process)

Let V be a finite-dimensional inner product space, with a given basis $\mathcal{B} = \{b_1, b_2, \dots b_n\}$.

A basis $\mathcal{B}$ is called orthogonal if its elements are all orthogonal to each other.

We can adjust $\mathcal{B}$ to an orthogonal basis $\mathcal{B}' = \{v_1, \dots, v_n\}$ as follows.

1. Write $v_1 = b_1$.
2. Write $v_2 = b_2 - \text{proj}_{v_1}(b_2) = b_2 - \dfrac{\langle v_1, b_2 \rangle}{||v_1||^2} v_1$.

   Then the pairs b_1, b_2 and v_1, v_2 span the same space, and $v_1 \perp v_2$.
3. Write $v_3 = b_3 - \text{proj}_{v_1}(b_3) - \text{proj}_{v_2}(b_3)$.

   Then the sets v_1, v_2, v_3 and b_1, b_2, b_3 span the same space, and v_3 is orthogonal to both v_1 and v_2.

   To see this note that

$$\langle v_1, v_3 \rangle = \langle v_1, b_3 \rangle - \frac{\langle v_1, b_3 \rangle}{\langle v_1, v_1 \rangle} \langle v_1, v_1 \rangle - c \cancel{\langle v_1, v_2 \rangle}$$

4. Continue in this way - at the kth step, form v_k by subtracting from b_k its projections on each of $v_1, \dots, v_n$.

Orthogonal projection on a subspace

The result of this process is a basis $\{v_1, \dots, v_n\}$ whose elements satisfy

$$\langle v_i, v_j \rangle = 0 \text{ for } i \neq j$$

We can adjust this basis to a orthonormal basis (consisting of orthogonal unit vectors) by replacing each v_i with its normalization $\hat{v}_i$.
From the Gram-Schmidt process, we have

Theorem

If V is a finite-dimensional inner product space, then V has an orthogonal (or orthonormal) basis.

Now let W be a subspace of V, and let $v \in V$. The orthogonal projection of v on W, denoted $\text{proj}_W(v)$, is defined to be the unique element u of W for which

$$v = u + v',$$

and $v' \perp w$ for all $w \in W$.

How to calculate a projection from an orthogonal basis

That $\text{proj}_W(v)$ exists follows from the fact that an orthogonal basis $\{b_1, \dots, b_k\}$ of W may be extended to an orthogonal basis $\mathcal{B} = \{b_1, \dots, b_k, c_{k+1}, \dots, c_n\}$ of W. Then v has a unique expression of the form

$$v = a_1 b_1 + \cdots + a_k b_k + a_{k+1} c_{k+1} + \cdots + a_n c_n, \text{ for scalars } a_i,$$

and $\text{proj}_W(v) = a_1 b_1 + \cdots + a_k b_k$.
Moreover, taking inner products with b_i gives $\langle v, b_i \rangle = a_i \langle b_i, b_i \rangle$, so that

$$\text{proj}_W(v) = \sum_{i=1}^{k} \frac{\langle v, b_i \rangle}{\langle b_i, b_i \rangle} b_i,$$

where $\{b_1, \dots, b_k\}$ is an orthogonal basis of W.

$\text{proj}_W(v)$ is the nearest point of W to v

Let $u = \text{proj}_W(v)$ and let w be any element of W. Then

$$\begin{aligned} d(v, w)^2 &= \langle v - w, v - w \rangle \\ &= \langle (v - u) + (u - w), (v - u) + (u - w) \rangle \\ &= \langle v - u, v - u \rangle + 2\langle v - u, u - w \rangle + \langle u - w, u - w \rangle \\ &= \langle v - u, v - u \rangle + \langle u - w, u - w \rangle \\ &\geq d(v, u)^2, \end{aligned}$$

with equality only if $w = \text{proj}_W(v)$.

Example In $\mathbb{R}^3$, find the unique point of the plane $x + 2y - z = 0$ that is nearest to the point $(1, 2, 2)$.

Application: least squares for overdetermined systems

Example Consider the following overdetermined linear system.

$$\begin{array}{rcrcr} 2x & + & y & = & 3 \\ x & - & y & = & 0 \\ x & - & 3y & = & -4 \end{array} \qquad \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix}$$

This system has three equations and only two variables. It is inconsistent and overdetermined - each pair of equations has a simultaneous solution, but all three do not.

Overdetermined systems arise quite often in applications, from observed data. Even if they do not have exact solutions, approximate solutions are of interest.

The least squares method

For a vector $b \in \mathbb{R}^3$, the system

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 1 & -3 \end{bmatrix}}_{A} \begin{bmatrix} x \\ y \end{bmatrix} = b$$

has a solution if and only if b belongs to the 2-dimensional linear span W of the columns of the coefficient matrix A: $v_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 1 \\ -1 \\ -3 \end{bmatrix}$.

If not, then the nearest element of W to B is $b' = \text{proj}_W(b)$, and our approximate solutions for x and y are the entries of the vector c in $\mathbb{R}^2$ for which $Ac = b'$. We know that $b' - b$ is orthogonal to v_1 and v_2, which are the rows of A^T. Hence

$$A^T(b' - b) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies A^T b' = A^T A c = A^T b \implies c = (A^T A)^{-1} A^T b$$

Example

In our example,

$$\begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix}$$

$$A = \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 1 & -3 \end{bmatrix}, \quad A^T = \begin{bmatrix} 2 & 1 & 1 \\ 1 & -1 & -3 \end{bmatrix}, \quad A^TA = \begin{bmatrix} 6 & -2 \\ -2 & 11 \end{bmatrix}, \quad A^Tb = \begin{bmatrix} 2 \\ 15 \end{bmatrix}.$$

The least squares solution is given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = c = (A^TA)^{-1}A^Tb = \frac{1}{62}\begin{bmatrix} 11 & 2 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} 2 \\ 15 \end{bmatrix} = \begin{bmatrix} \frac{26}{31} \\ \\ \frac{47}{31} \end{bmatrix}.$$