Orthogonal Projection, Least Squares, and the Pseudoinverse

N-20260305

§1.1 Projection onto a subspace

Let $U \subseteq \mathbb{R}^n$ have ordered basis $b_1, \ldots, b_m$ and put

$$B = (b_1, \ldots, b_m) \in \mathbb{R}^{n \times m}.$$

The orthogonal projection of x onto U has the form $B\lambda$. The error must be orthogonal to U:

$$B^T(x - B\lambda) = 0.$$

Thus

$$B^T B\lambda = B^T x, \qquad \lambda = (B^T B)^{-1} B^T x,$$

assuming the columns of B are independent. Therefore

$$\boxed{\pi_U(x) = B(B^T B)^{-1} B^T x.}$$

§1.2 Projection matrix and least squares

The matrix

$$P_U = B(B^T B)^{-1} B^T$$

is symmetric and idempotent:

$$P_U^T = P_U, \qquad P_U^2 = P_U.$$

For an overdetermined system $B\lambda \approx x$, minimizing $\|B\lambda - x\|^2$ gives the same normal equations. Least squares is orthogonal projection written in coordinates.

§1.3 Pseudoinverse

When B has full column rank, the Moore–Penrose pseudoinverse is

$$B^+ = (B^T B)^{-1} B^T.$$

For a general matrix $A = U\Sigma V^T$, the pseudoinverse is $A^+ = V\Sigma^+ U^T$, where nonzero singular values are inverted. This makes SVD the stable language of projection and least-squares computation.

§1.4 Projection geometry in the source diagram

The diagram is kept because projection and least squares are most transparent when the orthogonal error is visible. The vector being approximated is decomposed into its projection onto the column space and an error vector perpendicular to that space.

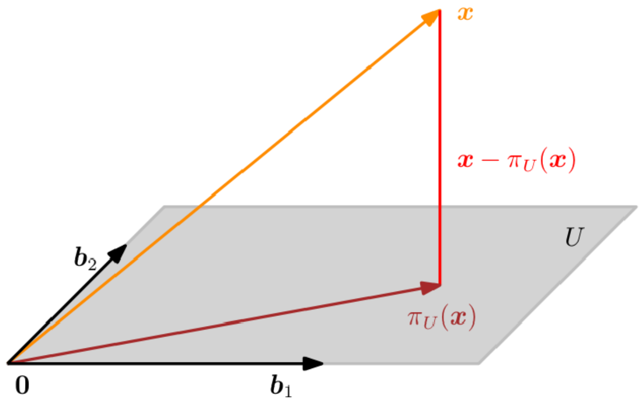

Figure 3.11 Projection onto a two-dimensional subspace U with basis $\boldsymbol{b}_1, \boldsymbol{b}_2$. The projection $\pi_U(\boldsymbol{x})$ of $\boldsymbol{x} \in \mathbb{R}^3$ onto U can be expressed as a linear combination of $\boldsymbol{b}_1, \boldsymbol{b}_2$ and the displacement vector $\boldsymbol{x} - \pi_U(\boldsymbol{x})$ is orthogonal to both $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$.

Figure 1.1: Projection onto a two-dimensional subspace: the residual $x - \pi_U(x)$ is orthogonal to the basis directions of U.

§1.5 Bridge to later ideas

Projection is the geometric heart of least squares: the best approximation is characterized not by making the error small in every coordinate, but by making the error orthogonal to every allowed correction direction.

§1.6 Projection theorem

Theorem 1.6.1 (Projection theorem)
For a closed subspace W of an inner-product space and a vector y, the best approximation $p \in W$ is characterized by

$$y - p \perp W.$$

For a closed subspace W of an inner-product space, the best approximation to y in W is the point $p \in W$ such that

$$y - p \perp W.$$

This orthogonality condition is the geometric core of least squares.

§1.7 Normal equations

For $Ax \approx b$, the residual $b - Ax$ must be orthogonal to the column space of A:

$$A^T(Ax - b) = 0.$$

Hence

$$A^TAx = A^Tb.$$

§1.8 Pseudoinverse and singular values

The Moore–Penrose pseudoinverse is most transparent through the singular value decomposition. If

$$A = U\Sigma V^T,$$

then

$$A^+ = V\Sigma^+ U^T.$$

Small singular values reveal unstable directions.

§1.9 Least squares as projection geometry

Least squares is projection geometry. Normal equations express orthogonality of residuals, the pseudoinverse selects the minimum-norm solution, and SVD exposes the stable and unstable directions of the problem.