Section 7.7: The projection theorem and its implications

Theorem 7.7.1 and Definition 7.7.2. *Let* $\mathbf{a} \in \mathbb{R}^n$ *be non-zero. Then every* $\mathbf{x} \in \mathbb{R}^n$ *can be expressed in exactly one way as* $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$*, where* $\mathbf{x}_1$ *is parallel to* $\mathbf{a}$ *and* $\mathbf{x}_2$ *is orthogonal to* $\mathbf{a}$*. We have*

$$\mathbf{x}_1 = \operatorname{proj}_{\mathbf{a}} \mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\,\mathbf{a} \qquad \textit{and} \qquad \mathbf{x}_2 = \mathbf{x} - \mathbf{x}_1.$$

$\operatorname{proj}_{\mathbf{a}} \mathbf{x}$ *is called the* ***orthogonal projection of*** $\mathbf{x}$ ***onto*** $\operatorname{span}\{\mathbf{a}\}$.

Theorem 7.7.3. *If we define an operator* $T\colon \mathbb{R}^n \to \mathbb{R}^n$ *by* $T(\mathbf{x}) = \operatorname{proj}_{\mathbf{a}} \mathbf{x}$*, then*

$$[T] = P = \frac{1}{\mathbf{a}^T\mathbf{a}}\,\mathbf{a}\,\mathbf{a}^T.$$

This matrix is symmetric ($P = P^T$) *and* ***idempotent*** ($P^2 = P$) *and has rank* 1*. If* $\mathbf{a}$ *is replaced by* $k\mathbf{a}$*,* P *does not change. If* $\mathbf{a} = \mathbf{u}$ *is a unit vector, then* $P = \mathbf{u}\,\mathbf{u}^T$*.*

Theorem 7.7.4. *Let* W *be a subspace of* $\mathbb{R}^n$*. Then every* $\mathbf{x} \in \mathbb{R}^n$ *can be expressed in exactly one way as* $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$*, where* $\mathbf{x}_1$ *is in* W *and* $\mathbf{x}_2$ *is in* $W^\perp$*.*

$\mathbf{x}_1$ is called the **orthogonal projection of $\mathbf{x}$ onto W** and is denoted $\operatorname{proj}_W \mathbf{x}$.

Theorem 7.7.5 and part of 7.7.6. *If* M *is a matrix whose columns form a basis for* W*, then* M^TM *is invertible and*

$$\operatorname{proj}_W \mathbf{x} = M\,(M^TM)^{-1}M^T\mathbf{x}.$$

If we define an operator $T\colon \mathbb{R}^n \to \mathbb{R}^n$ *by* $T(\mathbf{x}) = \operatorname{proj}_W \mathbf{x}$*, then*

$$[T] = P = M\,(M^TM)^{-1}M^T.$$

This matrix is symmetric and idempotent and has rank equal to the dimension of W*.*

Theorem 7.7.6. *If* P *is an* $n \times n$ *symmetric, idempotent matrix, then* T_P *is the orthogonal projection onto the subspace* $\operatorname{col}(P)$*.*

Section 7.8: Best approximations and least squares

Theorem 7.8.1 (Best approximation). *If* W *is a subspace of* $\mathbb{R}^n$ *and* $\mathbf{b}$ *is a point in* $\mathbb{R}^n$*, then* $\hat{\mathbf{w}} = \operatorname{proj}_W \mathbf{b}$ *is the unique* ***best approximation*** *to* $\mathbf{b}$ *from* W*. That is, for any other* $\mathbf{w}$ *in* W*,* $\|\mathbf{b} - \hat{\mathbf{w}}\| < \|\mathbf{b} - \mathbf{w}\|$*.*

Definition 7.8.2. Let A be an $m \times n$ matrix and $\mathbf{b}$ be in $\mathbb{R}^m$. A vector $\mathbf{x} \in \mathbb{R}^n$ is a **least squares solution** to $A\mathbf{x} = \mathbf{b}$ if it minimizes the error $\|\mathbf{b} - A\mathbf{x}\|$. The vector $\mathbf{b} - A\mathbf{x}$ is called the **least squares error vector** and the scalar $\|\mathbf{b} - A\mathbf{x}\|$ is called the **least squares error**.

The **normal system** associated to $A\mathbf{x} = \mathbf{b}$ is the system

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

Theorem 7.8.3. (*a*) *The* least squares *solutions of* $A\mathbf{x} = \mathbf{b}$ *are the* exact *solutions of the normal system.*

(*b*) *If* A *has full column rank (that is, the columns are linearly independent), then* A^TA *is invertible and the unique solution of the normal system is*

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}.$$

Theorem 7.8.4. $\hat{\mathbf{x}}$ *is a least squares solution to* $A\mathbf{x} = \mathbf{b}$ *if and only if the error vector* $\mathbf{b} - A\mathbf{x}$ *is orthogonal to* $\operatorname{col}(A)$*.*

Fitting a curve to data. Given data points (x_1, y_1), (x_2, y_2), …, (x_n, y_n) that are supposed to line on a line $y = a + bx$, we have

$$M\mathbf{v} = \mathbf{y}, \quad \text{where} \quad M = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

If the data do not exactly lie on a straight line, this system will have no solution, so we solve for a and b using the normal system

$$M^T M\mathbf{v} = M^T\mathbf{y}.$$

The line $y = a + bx$ is called the **least squares line of best fit**. Note that

$$M^T M = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \text{ and } M^T\mathbf{y} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}.$$

Moreover, $M^T M$ is invertible unless all of the x_i's are equal.

Section 7.9: Orthonormal bases and the Gram–Schmidt process

Theorem 7.9.1. *An orthogonal set of non-zero vectors is linearly independent.*

Theorem 7.9.2 and 7.9.4. *If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is an orthogonal basis for a subspace W, then*

$$\operatorname{proj}_W \mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{v}_1}{\|\mathbf{v}_1^2\|}\mathbf{v}_1 + \cdots + \frac{\mathbf{x} \cdot \mathbf{v}_k}{\|\mathbf{v}_k^2\|}\mathbf{v}_k.$$

If the vectors are orthonormal, then the denominators can be omitted. If $\mathbf{x}$ is in W, then $\operatorname{proj}_W \mathbf{x} = \mathbf{x}$*, so this gives a formula expressing $\mathbf{x}$ in terms of the given basis.*

Finding an orthonormal basis: Gram–Schmidt.

Theorem 7.9.5. *Every subspace of $\mathbb{R}^n$ has an orthonormal basis.*

Method: Start with any basis $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ for W. We will first find an orthogonal basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$:

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{w}_1 \\ \mathbf{v}_2 &= \mathbf{w}_2 - \frac{\mathbf{w}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\,\mathbf{v}_1 \\ \mathbf{v}_3 &= \mathbf{w}_3 - \frac{\mathbf{w}_3 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\,\mathbf{v}_1 - \frac{\mathbf{w}_3 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\,\mathbf{v}_2 \\ &\;\vdots \end{aligned}$$

Then set $\mathbf{q}_i = \dfrac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$ if an orthonormal basis is desired.

Theorem 7.9.7. *Let W be a subspace of $\mathbb{R}^n$. Every orthogonal (or orthonormal) set of vectors in W can be enlarged to an orthogonal (or orthonormal) basis for W.*