Linear Algebra: Sections 6.1 – 6.5

Objective: Introduce inner product, norm,

distance, orthogonality, orthonormal basis in $\mathbb{R}^n$

Inner product (or the dot product) of vectors in $\mathbb{R}^n$

Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, we have

$$\|\mathbf{u}\| = (u_1^2 + u_2^2)^{1/2},$$

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \{(u_1 - v_1)^2 + (u_2 - v_2)^2\}^{1/2},$$

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta,$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. We can do the same things in $\mathbb{R}^n$.

Definition The inner product of $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$ is defined by

$\mathbf{u} \cdot \mathbf{v} = \sum_{j=1}^{n} u_j v_j.$

Theorem 1 Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, and c be a scalar. Then

(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$,

(b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$.

(c) $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \dot{(} c\mathbf{v})$,

(d) $\mathbf{u} \cdot \mathbf{u} \geq 0$; $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = 0$.

Proof. Direct checking.

Definition The length (or norm) of $u \in \mathbb{R}^n$ is defined by $\|\mathbf{u}\| = (\mathbf{u} \cdot \mathbf{u})^{1/2}$.

The distance between $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$.

Theorem 2 (Pythagorean Theorem) Two vectors are orthogonal if and only if

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

Proof. Note that

$$\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})$$
$$= \mathbf{u}\cdot\mathbf{u} + \mathbf{v}\cdot\mathbf{v} + 2\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|^2 + 2\mathbf{u}\cdot\mathbf{v} + \|\mathbf{v}\|^2.$$

□

Definition Let $W \subseteq \mathbb{R}^n$ be a subspace. A vector $\mathbf{u} \in \mathbb{R}^n$ is orthogonal to W if is it orthogonal to every vector in W, and we write $\mathbf{u} \in W^\perp$.

The set $W^\perp = \{\mathbf{u} \in \mathbb{R}^n : \mathbf{u}\cdot\mathbf{v} \text{ for all } \mathbf{v} \in W\}$ is called the orthogonal complement of W.

Remark If $W \subseteq \mathbb{R}^n$ has a spanning set $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$, then $W^\perp = \operatorname{Nul} A^T$ with $A = [\mathbf{w}_1 \cdots \mathbf{w}_k]$.

Orthogonal sets

Definition A set $\{\mathbf{u}_1, \dots, \mathbf{u}_p\} \subseteq \mathbb{R}^n$ is orthogonal if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for all $i \neq j$. If in addition that $\|u_i\| = 1$ for all i, then it is an orthonormal set.

Remark Let $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\} \subseteq \mathbb{R}^n$ and $A = [\mathbf{u}_1 \cdots \mathbf{u}_p]$.
Then S is orthogonal if and only if $A^t A$ is a diagonal matrix.
The set S is orthonormal if and only if $A^t A = I_p$.

Theorem 4/5 An orthogonal set $S = \{\mathbf{u}_1, \dots, \mathbf{u}_p\} \subseteq \mathbb{R}^n$ of nonzero vectors is linearly independent.
If W is the span of $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$, then for every $\mathbf{u} \in W$ can be uniquely expressed as

$$\mathbf{u} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p \quad \text{with} \quad c_i = (\mathbf{u}_i \cdot \mathbf{u})/(\mathbf{u}_i \cdot \mathbf{u}_i) \qquad \text{for } i = 1, \dots, p.$$

Proof. Assume S is orthogonal with nonzero vectors. Let $A = [\mathbf{u}_1 \cdots \mathbf{u}_p]$. If $Ax = 0 \in \mathbb{R}^n$ for some $x \in \mathbb{R}^p$, then $A^t Ax = A^t 0 = 0 \in \mathbb{R}^p$. Note that $A^t A = D$ is a diagonal matrix with nonzero entries $\|\mathbf{u}_1\|^2, \dots, \|\mathbf{u}_p\|^2$. So, $Dx = 0$ implies $x = 0$. Hence the columns of A are linearly independent.

For the second statement, let

$$\mathbf{u} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p.$$

Then $c_i(\mathbf{u}_i \cdot \mathbf{u}_i) = \mathbf{u}_i \cdot \mathbf{u}$ so that $c_i = (\mathbf{u}_i \cdot \mathbf{u})/(\mathbf{u}_i \cdot \mathbf{u}_i)$ for every $i = 1, \dots, p$. □

Orthogonal projections

Orthogonal projection Suppose $\mathbf{u} \in \mathbb{R}^n$ is nonzero, and $\mathbf{y} \in \mathbb{R}^n$. The projection of $\mathbf{y}$ onto the line passing through $\mathbf{u}$ is: $\gamma\mathbf{u}$.

We have $\mathbf{y} - \gamma\mathbf{u} \in \{\mathbf{u}\}^\perp$.

Thus, $0 = (\mathbf{y} - \gamma\mathbf{u}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - \gamma\mathbf{u} \cdot \mathbf{u}$ with $\gamma = (\mathbf{y} \cdot \mathbf{u})/(\mathbf{u} \cdot \mathbf{u})$.

Hence,

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z},$$

where $\hat{\mathbf{y}} = \gamma\mathbf{u} = \text{proj}_L(\mathbf{y})$ is the orthogonal projection of $\mathbf{y}$ onto $L = \text{span}\{\mathbf{u}\}$.

Definition A set $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is orthonormal if it is an orthogonal set such that $\|\mathbf{u}_i\| = 1$ for all $i = 1, \ldots, p$.

Theorem 6/7 Let U be $m \times n$. Then U has orthonormal columns if and only if $U^T U = I_p$. In such a case, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

(a) $\|U\mathbf{x}\| = \|\mathbf{x}\|$,

(b) $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$,

(c) $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$.

Recall The orthogonal projection of $\mathbf{y}$ onto the line passing through the nonzero vector $\mathbf{u} \in \mathbb{R}^n$ is $\hat{\mathbf{y}} = \gamma \mathbf{u}$ where $\gamma = (\mathbf{y} \cdot \mathbf{u})/(\mathbf{u} \cdot \mathbf{u})$.

Theorem 8/9/10 Let $\mathbf{y} \in \mathbb{R}^n$ and W is a subspace of $\mathbb{R}^n$. Then $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ for unique $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. In particular, if W has an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, then

$$\hat{\mathbf{y}} = \gamma_1 \mathbf{u}_1 + \cdots + \gamma_p \mathbf{u}_p$$

with $\gamma_j = (\mathbf{u}_j \cdot \mathbf{y})/(\mathbf{u}_j \cdot \mathbf{u}_j)$. In particular, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis, then

$$\hat{\mathbf{y}} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p = UU^T\mathbf{y},$$

where $U = [\mathbf{u}_1 \cdots \mathbf{u}_p]$. Furthermore,

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \quad \text{for all other } \mathbf{v} \in W.$$

[Thus, $\hat{\mathbf{y}}$ is the best approximation of $\mathbf{y}$ by element in W.]

Proof. Note that $\hat{\mathbf{y}} \in W$ so that $\hat{\mathbf{y}} = \gamma_1 \mathbf{u}_1 + \cdots + \gamma_p \mathbf{u}_p$ for some scalars $\gamma_1, \ldots, \gamma_p$. As $\mathbf{u}_j \cdot \mathbf{y} = \gamma_j(\mathbf{u}_j \cdot \mathbf{u}_j)$,

$$\gamma_j = (\mathbf{u}_j \cdot \mathbf{y})/(\mathbf{u}_j \cdot \mathbf{u}_j) \quad \text{for } j = 1, \ldots, p.$$

We see that $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} \in W^\perp$ because

$$\mathbf{u}_j \cdot \mathbf{z} = \mathbf{u}_j \cdot (\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{u}_j \cdot \mathbf{y} - \mathbf{u}_j \cdot \hat{\mathbf{y}} = \gamma_j - \gamma_j = 0 \quad \text{for } j = 1, \ldots, p.$$

To prove that uniqueness of $\hat{\mathbf{y}}$ and $\mathbf{z}$,
suppose $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$.
Then $\mathbf{v} = \hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z} \in W \cap W^\perp$.
So, $\mathbf{v} \cdot \mathbf{v} = 0$ implies that $0 = \hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z}$.
Note that for any $\mathbf{v} \in W$ not equal to $\hat{\mathbf{y}}$, we have $(\hat{\mathbf{y}} - \mathbf{v}) \cdot (\mathbf{y} - \hat{\mathbf{y}}) = 0$ so that

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2 = \|\mathbf{y} - \mathbf{v}\|^2$$

as $\mathbf{y} - \hat{\mathbf{y}} \in W^\perp$ and $\hat{\mathbf{y}} - \mathbf{v} \in W$.
The asserted inequality follows. □

Finding an orthonormal basis

Question Can we find an orthogonal/orthonormal basis for a subspace of $\mathbb{R}^n$?

Theorem 11 (The Gram-Schmidt process.) Let $\{x_1, \ldots, x_p\}$ be a basis for W. We can construct an orthogonal basis as follows.

$$\mathbf{v}_1 = \mathbf{x}_1, \quad \mathbf{v}_2 = \gamma_2(\mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1),$$

$$\mathbf{v}_3 = \gamma_3(\mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2), \quad \cdots,$$

$$\mathbf{v}_p = \gamma_p(\mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}).$$

Here $\gamma_i > 0$ are chosen to make the entries of v_p easy to handle.

In particular, $\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \mathrm{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ for $k = 1, \ldots, p$.

One can normalize the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ to get an orthonormal basis:

$$\{\mathbf{v}_1/\|\mathbf{v}_1\|, \ldots, \mathbf{v}_p/\|\mathbf{v}_p\|\}.$$

QR Decomposition and Gram-Schmidt Process

QR Decomposition Let A be $m \times n$ with linearly independent columns.

Then $A = QR$ such that Q is an $m \times n$ matrix with orthonormal columns,

i.e., $Q^T Q = I_m$. and R is in upper triangular form and is nonsingular.

Algorithm. Let A be $n \times p$.

Apply Gram Schmidt procedure to the columns of A to get orthogonal columns $v_1, \ldots, v_p$.

Normalize $v_1, \ldots, v_p$ and put the resulting vectors in Q.

Set $Q^t A = R$.

Proof of Algorithm. Suppose A has columns $a_1, \ldots, a_p$, and Q has columns $u_1, \ldots, u_p$. From the Gram-Schmidt process:

$a_1 = r_{11} v_1$, $a_2 = r_{12} v_1 + r_{12} v_2$, etc.

So, $A = QR$ for an upper triangular matrix R.

Now, $Q^T Q = I_p$. Thus, $Q^T A = Q^T (QR) = R$. □

Let A be $m \times n$ and $\mathbf{b} \in \mathbb{R}^m$. If $A\mathbf{x} = \mathbf{b}$ has no solution, we determine $\hat{\mathbf{x}}$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

Theorem 13 The set of least square solutions is non-empty and equals the solution set of $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

Proof. We want to find $\hat{\mathbf{x}}$ in $W = \mathrm{Col}(A)$ that is nearest to $\mathbf{b}$.

Thus, $\mathbf{b} = \hat{\mathbf{x}} + \mathbf{z}$ such that $\mathbf{z} \in W^{\perp}$.

So, $\mathbf{b} - \hat{\mathbf{x}} \in W^{\perp}$, i.e., $0 = A^T(\mathbf{b} - \hat{\mathbf{x}}) = A^T\mathbf{b} - A^T A\hat{\mathbf{x}}$. □

More results

Theorem 14/15 Let A be an $m \times n$ matrix. The following are equivalent.

(a) The equation $Ax = b$ has a unique least square solution.

(b) The columns of A are linearly independent.

(c) The matrix $A^T A$ is invertible.

In such a case $\hat{x} = (A^T A)^{-1} A^T b$ is the unique solution.

Furthermore, if $A = QR$ with $Q^T Q = I_m$, then $\hat{x} = R^{-1} Q^T b$ and $R\hat{x} = Q^T b$.

Applications

Example If $(x_1, y_1), \dots, (x_m, y_m)$ are given, we want to find a line $y = a_0 + a_1 x$ to best approximate the data, we compute

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

So, we want to solve

$$\begin{bmatrix} m & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}.$$

If we want to find a quadratic curve $y = a_0 + a_1 x + a_2 x^2$ to best approximate the data, we compute

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$