17 Lecture 17: Diagonalization of Small Matrices

Recall **eigenvalues and eigenvectors:** Let $\mathbf{A}$ be an $n \times n$ matrix,

- **An eigenvalue of A** is a real (or complex) scalar λ such that equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ has nontrivial solution. Equivalently, λ is an eigenvalue if and only if it is a **root of the characteristic polynomial**
$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}),$$
or, solution of the **characteristic equation**
$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0.$$

- **An eigenvector v** (corresponding with an eigenvalue λ) of the matrix $\mathbf{A}$ is a nontrivial solution of equation
$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$$

17.1 Eigenspace

The eigenspace corresponding with an eigenvalue λ of an $n \times n$ matrix $\mathbf{A}$ is the set of all eigenvectors corresponding with λ. Equivalently, eigenspace of $\mathbf{A}$ corresponding with an eigenvalue λ is the null space of the matrix $\mathbf{A} - \lambda\mathbf{I}$.

Thus, finding eigenvectors corresponding with eigenvalue of matrix $\mathbf{A}$ is equivalent to finding the basis of $\mathrm{Nul}(\mathbf{A} - \lambda\mathbf{I})$.

Example 17.1. *Find eigenvalues and associated eigenvectors of the matrix*
$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

Solution: (To find eigenvalues:) The characteristic polynomial
$$p_{\mathbf{A}}(\lambda) = \begin{vmatrix} 1-\lambda & 2 & 3 \\ 2 & 4-\lambda & 6 \\ 3 & 6 & 9-\lambda \end{vmatrix} = -\lambda^3 + 14\lambda^2$$

$p_{\mathbf{A}}(\lambda) = 0$ if and only if $\lambda = 0$, or $\lambda = 14$. Thus, matrix $\mathbf{A}$ has three eigenvalues (two distinct eigenvalues): $\lambda_1 = \lambda_2 = 0$ (multiplicity 2) and $\lambda_3 = 14$.

(To find eigenvectors:) For $\lambda_1 = 0$, $\mathbf{A} - \lambda_1\mathbf{I} = \mathbf{A}$. Row reducing the augmented matrix for the linear system $(\mathbf{A} - \lambda_1)\mathbf{x} = \mathbf{0}$ gives
$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 4 & 6 & 0 \\ 3 & 6 & 9 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Solution to the linear system is

$$x_1 = -2x_2 - 3x_3; \quad x_2, x_3 \text{ are free.}$$

The parametric vector form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_2 - 3x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}.$$

Thus, $\mathbf{A}$ has two linearly independent eigenvectors associated with $\lambda = 0$, which are $\mathbf{v}_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$. Equivalently, a basis for the eigenspace associated with $\lambda = 0$ is $\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}$.

For $\lambda_3 = 14$, $\mathbf{A} - \lambda_2 \mathbf{I} = \begin{bmatrix} -13 & 2 & 3 \\ 2 & -10 & 6 \\ 3 & 6 & -5 \end{bmatrix}$. Row reducing the augmented matrix for the linear system $(\mathbf{A} - \lambda_2 \mathbf{I}) = \mathbf{0}$ gives

$$\left[\begin{array}{ccc|c} -13 & 2 & 3 & 0 \\ 2 & -10 & 6 & 0 \\ 3 & 6 & -5 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -1/3 & 0 \\ 0 & 1 & -2/3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The solution is $x_1 = x_1/3, \quad x_2 = 2x_3/3, \quad x_3$ is free. The parametric vector form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3/3 \\ x_3/3 \\ x_3 \end{bmatrix} = \frac{x_3}{3} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Thus, an eigenvector is $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Equivalently, a basis for the eigenspace associated with the eigenvalue $\lambda_3 = 14$ is $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\}$.

17.2 Diagonalization

Recall: A diagonalization of an $n \times n$ matrix $\mathbf{A}$ is a factorization of the form

$$\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1},$$

where, $\mathbf{S}$ is a nonsingular matrix and $\mathbf{D}$ is a diagonal matrix.

The theorem below holds for a square matrix of the general size:

Theorem 9. *An $n \times n$ matrix $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ has n linearly independent eigenvectors. In this case,*

$$\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1},$$

where columns of $\mathbf{S}$ are the n eigenvectors and the diagonal matrix $\mathbf{D}$ is formed eigenvalues of $\mathbf{A}$ that correspond to the eigenvector in $\mathbf{S}$. That is,

$$\mathbf{S} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix},$$

where $\mathbf{A}\mathbf{v}_\ell = \lambda_\ell \mathbf{v}_\ell, \quad \ell = 1, 2, \dots, n.$

The form $\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}$ is called the diagonalization of $\mathbf{A}$.

Example 17.2. *Find the diagonalization of* $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$.

Solution: From Example 17.1, matrix $\mathbf{A}$ has three eigenvalues $\lambda_{1,2} = 0$, and $\lambda_3 = 14$ and the corresponding eigenvectors $\mathbf{v}_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ (respectively). The diagonalization of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1},$$

where

$$\mathbf{S} = \begin{bmatrix} -2 & -3 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix}, \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 14 \end{bmatrix}.$$

Example 17.3. *Find the diagonalization of the following matrix (if possible)*

$$\mathbf{A} = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$$

Solution: *We first find the eigenvalues and the corresponding eigenvectors for matrix* $\mathbf{A}$*. The characteristic polynomial*

$$p_{\mathbf{A}}(\lambda) = \begin{vmatrix} 4-\lambda & -1 & 6 \\ 2 & 1-\lambda & 6 \\ 2 & -1 & 8-\lambda \end{vmatrix} = -\lambda^3 + 13\lambda^2 - 40\lambda + 36$$

$$p_{\mathbf{A}}(\lambda) = 0 \iff -\lambda^3 + 13\lambda^2 - 40\lambda + 36 = 0 \iff (\lambda-9)(\lambda-2)^2 = 0,$$

which implies $\lambda = 9$*, or* $\lambda = 2$*. So,* $\mathbf{A}$ *has three eigenvalues* $\lambda_1 = 9$*,* $\lambda_2 = \lambda_3 = 2$ *(two distinct eigenvalues).*

For $\lambda_1 = 9$*,* $\mathbf{A} - \lambda_1 \mathbf{I} = \begin{bmatrix} -5 & -1 & 6 \\ 2 & -8 & 6 \\ 2 & -1 & -1 \end{bmatrix}$*. Row reducing the augmented matrix of the linear system* $(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{x} = \mathbf{0}$ *gives,*

$$\begin{bmatrix} -5 & -1 & 6 & 0 \\ 2 & -8 & 6 & 0 \\ 2 & -1 & -1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The solution is $x_1 = x_3$*,* $x_2 = x_3$*,* x_3 *is free. The parametric vector form*

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Thus, there is an eigenvector associated with the eigenvalue $\lambda = 9$*, which is* $v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

For $\lambda_2 = 2$*,* $\mathbf{A} - \lambda_2 \mathbf{I} = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}$*. Row reducing the augmented matrix of the linear system* $(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{x} = \mathbf{0}$ *gives,*

$$\begin{bmatrix} 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1/2 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The solution is $x_1 = 1/2x_2 - 3x_3$*,* x_2 *and* x_3 *are free. The parametric vector form*

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_2/2 - 3x_3 \\ x_2 \\ x_3 \end{bmatrix} = \frac{x_2}{2} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}.$$

There are two eigenvectors associated with$\lambda = 2$*, which are* $v_2 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ *and* $v_3 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$. *The diagonalization of* $\mathbf{A}$ *is*

$$\mathbf{A} = \mathbf{SDS}^{-1},$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & -3 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

17.3 Real matrix with complex eigenvalues

If an $n \times n$ real matrix has a complex eigenvalue $\lambda = a + bi$ with a complex eigenvector $\mathbf{v}$, then $\overline{\lambda}$ is also an eigenvalue and the corresponding eigenvector is $\overline{\mathbf{v}}$.

Remark 17.1. *An* $n \times n$ *matrix has exactly* n *eigenvalues including complex values and multiplicity.*

Example 17.4. *Diagonalize the matrix* $\mathbf{A}$ *(if possible)*

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$$

Solution: We first find eigenvalues and the corresponding eigenvectors. The characteristic polynomial

$$p_{\mathbf{A}}(\lambda) = \begin{vmatrix} 2-\lambda & -1 \\ 1 & 2-\lambda \end{vmatrix} = (2-\lambda)^2 + 1$$

$p_{\mathbf{A}}(\lambda) = 0$ implies $\lambda_{1,2} = 2 \pm i$. Thus, $\mathbf{A}$ has complex eigenvalues $\lambda_1 = 2 + i$ and $\lambda_2 = 2 - i = \overline{\lambda_1}$.

For $\lambda_1 = 2 + i$, $\mathbf{A} - \lambda_1 \mathbf{I} = \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix}$. Row reducing the augmented matrix of the linear system $\begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ gives,

$$\begin{bmatrix} -i & -1 & 0 \\ 1 & -i & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -i & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The solution is $x_1 = ix_2$, x_2 is free. The parametric vector form $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ix_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} i \\ 1 \end{bmatrix}$. The eigenvector associated with eigenvalue $\lambda_1 = 2 + i$ is $\mathbf{v}_1 = \begin{bmatrix} i \\ 1 \end{bmatrix}$. Thus, the eigenvector associated with eigenvalue $\lambda_2 = 2 - i$ is $\mathbf{v}_2 = \overline{\mathbf{v}_1} = \begin{bmatrix} -i \\ 1 \end{bmatrix}$. The diagonalization of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1},$$

where

$$\mathbf{S} = \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 2+i & 0 \\ 0 & 2-i \end{bmatrix}.$$

18 Lecture 18: Orthogonal Projection, Least Squares Solutions and Regression

18.1 Orthogonal, orthogonal complements

1. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbf{R}^n$ are orthogonal (to each other) is $\mathbf{u} \cdot \mathbf{v} = 0$.

2. A vector $\mathbf{u} \in \mathbf{R}^n$ is **orthogonal to a subspace** $V \subset \mathbf{R}^n$ if $\mathbf{u}$ is orthogonal to all vectors $\mathbf{v} \in V$.

3. The set of all vectors $\mathbf{u} \in \mathbf{R}^n$ that is orthogonal to a subspace V of $\mathbf{R}^n$ is a subspace of $\mathbf{R}^n$. This subspace is called the **orthogonal complement of** V in $\mathbf{R}^n$ and denoted by V^T.

4. If V is a subspace of $\mathbf{R}^n$, $V \cap V^T = \{\mathbf{0}\}$ and $V \cup V^T = \mathbf{R}^n$. As a result, $\dim(V) + \dim(V^T) = n$.

5. Example: If $\mathbf{A}$ is an $m \times n$ real matrix,
   - $(\text{Row}(\mathbf{A}))^T = \text{Nul}(\mathbf{A})$ and they are subspace of $\mathbf{R}^n$.
   - $(\text{Col}(\mathbf{A}))^T = \text{Nul}(\mathbf{A}^T)$ and they are subspace of $\mathbf{R}^m$.

18.2 Orthogonal projection

Assume that V is a subspace of $\mathbf{R}^n$ and $\mathbf{x}$ is a vector in $\mathbf{R}^n$. The **orthogonal projection** of $\mathbf{x}$ into V is the vector $\hat{\mathbf{x}} \in V$ such that the distance between $\mathbf{x}$ and $\hat{\mathbf{x}}$, i.e., $\|\mathbf{x} - \hat{\mathbf{x}}\|$, is smallest. From the Pythapogean theorem, distance between $\mathbf{x}$ and $\hat{\mathbf{x}}$ is smallest if and only if $\mathbf{x} - \hat{\mathbf{x}}$ is orthogonal to V.

The vector $\hat{\mathbf{x}} \in \mathbf{V}$ such that $\mathbf{x} - \hat{\mathbf{x}}$ is orthogonal to V is called the **orthogonal projection of x onto** V, and written "**proj**$_V$**x**". Obviously, vector $\mathbf{x} - \hat{\mathbf{x}}$ is in the orthogonal complement of $\mathbf{V}$. In this case, vector $\mathbf{x}$ is written as

$$\mathbf{x} = \hat{\mathbf{x}} + (\mathbf{x} - \hat{\mathbf{x}}), \tag{18.1}$$

where $\hat{\mathbf{x}} \in V$ and $\mathbf{x} - \hat{\mathbf{x}} \in \mathbf{V}^T$. The sum (18.1) is called the orthogonal decomposition of $\mathbf{x}$.

18.2.1 Calculation of the orthogonal projection of x onto V:

Assume that $S = \{v_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is a basis for V. We denote the matrix $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_p \end{bmatrix}$, whose columns are formed by S. We find a matrix $\mathbf{P}$ such that $\hat{\mathbf{x}} = \mathbf{P}\mathbf{x}$, where $\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{x}$ onto V. Matrix $\mathbf{P}$ must depend on $\mathbf{A}$.

To find $\mathbf{P}$, we first see that $\hat{\mathbf{x}} \in V$ is written as a linear combination of the basis S as

$$\hat{\mathbf{x}} = \alpha_1 \mathbf{v}_1 + \ldots \alpha_p \mathbf{v}_p = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_p \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} = \mathbf{A}\boldsymbol{\alpha}, \quad \text{where } \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix}.$$

In addition, $\mathbf{x} - \hat{\mathbf{x}}$ is orthogonal to V, thus $\mathbf{x} - \hat{\mathbf{x}}$ is orthogonal to all vector in its basis, which implies

$$\mathbf{v}_j \cdot (\mathbf{x} - \hat{\mathbf{x}}) = 0, \quad \text{for all } \ j = 1, 2, \ldots p$$

This system is written as

$$\mathbf{A}^T(\mathbf{x} - \hat{\mathbf{x}}) = \mathbf{0}, \quad \text{or} \quad \mathbf{A}^T\mathbf{A}\alpha = \mathbf{A}^T\mathbf{x}.$$

Since columns of $\mathbf{A}$ are linearly independent, the matrix $\mathbf{A}^T\mathbf{A}$ is invertible. So, equation $\mathbf{A}^T\mathbf{A}\alpha = \mathbf{A}^T\mathbf{x}$ has a unique solution $\boldsymbol{\alpha}$, which is $\boldsymbol{\alpha} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x}$. Plugging it into the equation $\hat{\mathbf{x}} = \mathbf{A}\boldsymbol{\alpha}$, we end up with

$$\hat{\mathbf{x}} = \mathbf{A}\boldsymbol{\alpha} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x} = \mathbf{P}\mathbf{x}, \quad \text{with } \mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T.$$

Matrix $\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the matrix of the **orthogonal projection** of the orthogonal projection onto V.

Definition 18.1. *A projection matrix that is symmetric is called an* **orthogonal projection matrix**.

Note: Matrix $\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is an orthogonal projection matrix.

Indeed, $\mathbf{P}^T = \mathbf{P}$ and $\mathbf{P}^2 = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{P}$.

18.2.2 Orthogonal sets

A set of p vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ in $\mathbf{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set are orthogonal, that is if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$.

Example 18.1. *The set of three vectors* $\left\{ \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 7 \\ -4 \\ -1 \end{bmatrix} \right\}$ *is an orthogonal set.*

- If S is an orthogonal set of nonzero vectors in $\mathbf{R}^n$, then S is a linearly independent set.
- (Definition) An **orthogonal basis** of a subspace V of $\mathbf{R}^n$ is a basis for V that is also an orthogonal set.
- If $\mathbf{A}$ is an $m \times n$ real matrix and columns of $\mathbf{A}$ form an orthogonal set, then the Gram matrix $\mathbf{A}^T\mathbf{A}$ is an $n \times n$ diagonal matrix.

18.2.3 Orthonormal sets

A set of p vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ in $\mathbf{R}^n$ is said to be an **orthonormal set** if S is an orthogonal set of unit vector, i.e., $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$ and $\|u_j\| = 1$ for all $j = 1, 2, \ldots, p$.

Example 18.2. *The set of three vectors* $\left\{ \frac{1}{\sqrt{11}} \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \frac{1}{\sqrt{66}} \begin{bmatrix} 7 \\ -4 \\ -1 \end{bmatrix} \right\}$ *is an orthonormal set.*

- (Definition) An **orthonormal basis** of a subspace V of $\mathbf{R}^n$ is a basis for V that is also an orthonormal set.
- If $\mathbf{A}$ is an $m \times n$ real matrix, columns of $\mathbf{A}$ form an orthonormal set (also say, $\mathbf{A}$ has orthonormal columns) if and only if the Gram matrix $\mathbf{A}^T\mathbf{A}$ is an $n \times n$ identity matrix.
- (Properties) Let $\mathbf{A}$ be an $m \times n$ matrix with orthonormal columns. Then

1. $\|\mathbf{Ax}\| = \|\mathbf{x}\|$
2. $(\mathbf{Ax}) \cdot (\mathbf{Ay}) = \mathbf{x} \cdot \ \mathbf{y}$
3. $(\mathbf{Ax}) \cdot (\mathbf{Ay}) = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$.

- If columns of $\mathbf{A}$ is an orthonormal basis for a subspace V in $\mathbf{R}^n$, then the orthogonal projection matrix onto V is simplified as
$$\mathbf{P} = \mathbf{AA}^T$$
(due to $\mathbf{A}^T\mathbf{A} = \mathbf{I}_n$).

18.2.4 Projection onto the orthogonal complement

Recall the orthogonal decomposition
$$\mathbf{x} = \hat{\mathbf{x}} + (\mathbf{x} - \hat{\mathbf{x}}),$$
where $\hat{\mathbf{x}} = \mathbf{Px} \in V$ the orthogonal projection of $\mathbf{x}$ onto V. Then, $\mathbf{x} - \hat{\mathbf{x}}$ is in V^T, the orthogonal complement of V. We see that
$$x - \hat{\mathbf{x}} = \mathbf{x} - \mathbf{Px} = (\mathbf{I} - \mathbf{P})\mathbf{x}$$
Matrix $\mathbf{Q} := \mathbf{I} - \mathbf{P}$ is the orthogonal projection matrix onto the complement of V.

18.3 Least square solution

A linear system $\mathbf{Ax} = \mathbf{b}$ is inconsistent if $\mathbf{b} \notin \mathrm{Col}(\mathbf{A})$. In this situations, it is often useful to consider $\mathbf{x} \in \mathbf{R}^n$ such that $\mathbf{Ax}$ is the best approximation of $\mathbf{b}$ in $\mathrm{Col}(\mathbf{A})$. That happens when $\mathbf{Ax}$ is the orthogonal projection, $\widehat{\mathbf{b}}$, of $\mathbf{b}$ onto $\mathrm{Col}(\mathbf{A})$. The solution to equation
$$\mathbf{Ax} = \widehat{\mathbf{b}}, \quad \text{where} \quad \widehat{\mathbf{b}} = \mathrm{proj}_{\mathrm{Col}(\mathbf{A})}\mathbf{b}$$
is called the **least-square solution** of the equation $\mathbf{Ax} = \mathbf{b}$.

Remark 18.1. *Equation* $\mathbf{Ax} = \widehat{\mathbf{b}}$ *is consistent due to* $\widehat{\mathbf{b}} = \mathrm{proj}_{\mathrm{Col}(\mathbf{A})}\mathbf{b}$ *is in* $\mathrm{Col}(\mathbf{A})$.

18.3.1 Solving for the least-square solution

$\mathbf{x} \in \mathbf{R}^n$ is a least-square solution of $\mathbf{Ax} = \mathbf{b}$ is and only if $\mathbf{b} - \mathbf{Ax}$ is orthogonal to the column space of $\mathbf{A}$. That is,
$$\mathbf{A}^T(\mathbf{b} - \mathbf{Ax}) = 0,$$
which implies
$$\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$$
The linear equation $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ is then solved by row reducing the augmented matrix $\begin{bmatrix}\mathbf{A}^T\mathbf{A} & \mathbf{A}^T\mathbf{b}\end{bmatrix}$.

- If columns of $\mathbf{A}$ are **linearly independent**, equation $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ has a unique solution, with is
$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$
- If columns of $\mathbf{A}$ are **linearly dependent**, equation $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ has a infinitely many solutions.

18.4 Linear Regression

Consider a collection of p points $(x_1, y_1), (x_2, y_2), \ldots, (x_p, y_p)$ in $\mathbf{R}^2$. The linear regression method is to find a best-fit line of the form $y = ax + b$ for these p points. To find the line, we aim to find coefficients a and b. From the data, we set the linear system for the model coefficients, which is

$$\begin{cases} ax_1 + b = y_1 \\ ax_2 + b = y_2 \\ \qquad \vdots \\ ax_p + b = y_p, \end{cases}$$

or equivalent to

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_p & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}.$$

This linear system is generally inconsistent. So, we find the least-square solution of this system. The first and the second components of the least-square solution are values for a and b (respectively).

Example 18.3. *Find values of the parameters a and b such that the line $y = ax + b$ best fits the following points: $(x_1, y_1) = (-6, -1)$, $(x_2, y_2) = (-2, 2)$, $(x_3, y_3) = (1, 1)$, $(x_4, y_4) = (7, 6)$, $(x_5, y_5) = (4, 8)$.*

Solution: The linear system for the model coefficients is

$$\begin{cases} -6a + b &= -1 \\ -2a + b &= 2 \\ a + b &= 1 \\ 7a + b &= 6 \\ 4a + b &= 8, \end{cases}$$

which is equivalent to

$$\begin{bmatrix} -6 & 1 \\ -2 & 1 \\ 1 & 1 \\ 7 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \\ 8 \end{bmatrix}.$$

We find the least-square solution, which solves

$$\begin{bmatrix} -6 & -2 & 1 & 7 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -6 & 1 \\ -2 & 1 \\ 1 & 1 \\ 7 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -6 & -2 & 1 & 7 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \\ 8 \end{bmatrix},$$

or

$$\begin{bmatrix} 106 & 4 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 77 \\ 16 \end{bmatrix}.$$

Matrix $\begin{bmatrix} 106 & 4 \\ 4 & 5 \end{bmatrix}$ is invertible, so the solution is

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{514} \begin{bmatrix} 5 & -4 \\ -4 & 106 \end{bmatrix} \begin{bmatrix} 77 \\ 16 \end{bmatrix} = \begin{bmatrix} 321/514 \\ 694/257 \end{bmatrix}.$$

Thus, the best-fit line is

$$y = \frac{321}{514}x + \frac{694}{257}.$$