MA110:
Lecture 14

Saurav Bhaumik
Department of Mathematics
IIT Bombay
Spring 2025

Review of the last lecture

- On $\mathbb{K}^{n\times 1}$, we have the inner product defined by
$$\langle \mathbf{x}, \mathbf{y}\rangle := \mathbf{x}^*\mathbf{y} = \overline{x}_1 y_1 + \cdots + \overline{x}_n y_n.$$
It is positive definite, linear in the 2nd variable, conjugate symmetric (hence conjugate linear in the 1st variable).
- The norm of $\mathbf{x} \in \mathbb{K}^{n\times 1}$ is defined by
$$\|\mathbf{x}\| := \langle \mathbf{x}, \mathbf{x}\rangle^{1/2} = (|x_1|^2 + \cdots + |x_n|^2)^{1/2}.$$
It satisfies the triangle inequality: $\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$.
- The projection of a vector $\mathbf{x}$ in the direction of a nonzero vector $\mathbf{y}$ is a scalar multiple of $\mathbf{y}$ given by
$$P_{\mathbf{y}}(\mathbf{x}) := \frac{\langle \mathbf{y}, \mathbf{x}\rangle}{\langle \mathbf{y}, \mathbf{y}\rangle}\mathbf{y} \quad \text{and it satisfies} \quad \langle \mathbf{y}, \mathbf{x} - P_{\mathbf{y}}(\mathbf{x})\rangle = 0.$$
- A subset E of $\mathbb{K}^{n\times 1}$ if orthogonal if $\langle \mathbf{x}, \mathbf{y}\rangle = 0$ for all $\mathbf{x}, \mathbf{y} \in E$ with $\mathbf{x} \neq \mathbf{y}$. If E is orthogonal and $\mathbf{0} \notin E$, then E is linearly independent.

Gram-Schmidt Orthogonalization Process

The Gram-Schmidt Orthogonalization Process (G-S OP) transforms a linearly independent set in $\mathbb{K}^{n\times 1}$ to an orthogonal set without altering the span. It is given by the following.

Let $(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ be an ordered linearly independent set in $\mathbb{K}^{n\times 1}$. Define $\mathbf{y}_1 := \mathbf{x}_1$. Let $1 \le j < k$. Suppose we have found $\mathbf{y}_1, \ldots, \mathbf{y}_j$ in $\mathbb{K}^{n\times 1}$ such that the set $\{\mathbf{y}_1, \ldots, \mathbf{y}_j\}$ is orthogonal, and also $\operatorname{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_j\} = \operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_j\}$. Define

$$\mathbf{y}_{j+1} := \mathbf{x}_{j+1} - P_{\mathbf{y}_1}(\mathbf{x}_{j+1}) - \cdots - P_{\mathbf{y}_j}(\mathbf{x}_{j+1}).$$

Then $\operatorname{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_{j+1}\} = \operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_{j+1}\}$ since $\mathbf{y}_{j+1} \in \operatorname{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_j, \mathbf{x}_{j+1}\} = \operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_j, \mathbf{x}_{j+1}\}$ and $\mathbf{x}_{j+1} \in \operatorname{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_j, \mathbf{y}_{j+1}\}$.

To show that the set $\{\mathbf{y}_1, \ldots, \mathbf{y}_{j+1}\}$ is orthogonal, it is enough to show that $\mathbf{y}_{j+1} \in \{\mathbf{y}_1, \ldots, \mathbf{y}_j\}^{\perp}$.

Let $i \in \{1, \ldots, j\}$. Then

$$\begin{aligned}
\langle \mathbf{y}_i, \mathbf{y}_{j+1} \rangle &= \langle \mathbf{y}_i, \mathbf{x}_{j+1} - P_{\mathbf{y}_1}(\mathbf{x}_{j+1}) - \cdots - P_{\mathbf{y}_j}(\mathbf{x}_{j+1}) \rangle \\
&= \langle \mathbf{y}_i, \mathbf{x}_{j+1} \rangle - \langle \mathbf{y}_i, P_{\mathbf{y}_1}(\mathbf{x}_{j+1}) \rangle - \cdots - \langle \mathbf{y}_i, P_{\mathbf{y}_j}(\mathbf{x}_{j+1}) \rangle \\
&= \langle \mathbf{y}_i, \mathbf{x}_{j+1} \rangle - \langle \mathbf{y}_i, P_{\mathbf{y}_i}(\mathbf{x}_{j+1}) \rangle \quad (\text{since } \mathbf{y}_i \perp \mathbf{y}_j,\ i \neq j) \\
&= \langle \mathbf{y}_i, \mathbf{x}_{j+1} - P_{\mathbf{y}_i}(\mathbf{x}_{j+1}) \rangle \\
&= 0 \quad (\text{by the important property of the projection}).
\end{aligned}$$

We remark that since the set $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is linearly independent, all vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k$ constructed in the G-S OP are nonzero: Clearly, $\mathbf{y}_1 = \mathbf{x}_1 \neq \mathbf{0}$. Also, if $\mathbf{y}_{j+1} = \mathbf{0}$ for some $j \in \{1, \ldots, k-1\}$, then $\mathbf{x}_{j+1}$ would belong to $\operatorname{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_j\} = \operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_j\}$.

This completes the construction of the G-S OP.

Definition

An orthogonal set whose elements are unit vectors is called an **orthonormal set**.

Any orthogonal set whose elements are nonzero vectors can always be turned into an orthonormal set by dividing each element by its own norm.

Thus given an ordered linearly independent set $(\mathbf{x}_1, \ldots, \mathbf{x}_k)$, we can construct an ordered orthogonal set $(\mathbf{y}_1, \ldots, \mathbf{y}_k)$ by the G-S OP, and if we let $\mathbf{u}_j := \mathbf{y}_j / \|\mathbf{y}_j\|$ for $j = 1, \ldots, k$, then $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is an ordered orthonormal set such that $\operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} = \operatorname{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_k\} = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$.

Example: For $j = 1, \ldots, n$, let $\mathbf{x}_j := j(\mathbf{e}_1 + \cdots + \mathbf{e}_j)$. Then $E := (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is an ordered linearly independent subset of $\mathbb{K}^{n \times 1}$. We claim that the G-S OP gives $\mathbf{y}_j := j\mathbf{e}_j$ for $j = 1, \ldots, n$. Indeed, $\mathbf{y}_1 := \mathbf{x}_1 = \mathbf{e}_1$. Also, assuming that $\mathbf{y}_j = j\mathbf{e}_j$, we see that

$$
\begin{aligned}
\mathbf{y}_{j+1} &= \mathbf{x}_{j+1} - P_{\mathbf{y}_1}(\mathbf{x}_{j+1}) - \cdots - P_{\mathbf{y}_j}(\mathbf{x}_{j+1}) \\
&= (j+1)(\mathbf{e}_1 + \cdots + \mathbf{e}_{j+1}) - (j+1)\mathbf{e}_1 - \cdots - (j+1)\mathbf{e}_j \\
&= (j+1)\mathbf{e}_{j+1}.
\end{aligned}
$$

Hence our claim is justified. Since $\|\mathbf{y}_j\| = j$ for each j, we let $\mathbf{u}_j := \mathbf{y}_j/j$, so that $\mathbf{u}_j = \mathbf{e}_j$ for each $j = 1, \ldots, n$. Clearly, $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is an ordered orthonormal set in $\mathbb{K}^{n\times 1}$.

Definition

Let V be a subspace of $\mathbb{K}^{n\times 1}$. An **orthonormal basis** *for V is a basis for V which is an orthonormal subset of V.*

The G-S OP enables us to modify a given basis for a subspace of $\mathbb{K}^{n\times 1}$ to an orthonormal basis for that subspace.

Also, we can expand an orthonormal set in V to a possibly larger orthonormal set in V as follows.

Proposition

Let V be a subspace of $\mathbb{K}^{n\times 1}$, and let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be an orthonormal set in V. Then there is an orthonormal basis for V which contains $\mathbf{u}_1, \ldots, \mathbf{u}_k$.

Proof. If $\mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} = V$, then there is nothing to prove.

Now suppose $\mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \neq V$. Let $\dim V = r$. Since the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \neq V$ is linearly independent, there are $\mathbf{y}_{k+1}, \ldots, \mathbf{y}_r$ in V such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{y}_{k+1}, \ldots, \mathbf{y}_r\}$ is a basis for V. By the G-S OP, we can find $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_r$ in V such that the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_r\}$ is orthonormal and its span is equal to $\mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{y}_{k+1}, \ldots, \mathbf{y}_r\} = V$. $\square$

Corollary

Every nonzero vector subspace V has an orthonormal basis.

Proof. If $\mathbf{0} \neq \mathbf{x}_1 \in V$, then extend $\left\{\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}\right\}$ to an orthornormal basis, using the above proposition.

Example: Let W be the subspace of $\mathbb{K}^{4\times 1}$ spanned by the vectors $\mathbf{x}_1 := \begin{bmatrix}1 & 1 & 0 & 1\end{bmatrix}^\mathsf{T}$, $\mathbf{x}_2 := \begin{bmatrix}1 & -2 & 0 & 0\end{bmatrix}^\mathsf{T}$ and $\mathbf{x}_3 := \begin{bmatrix}1 & 0 & -1 & 2\end{bmatrix}^\mathsf{T}$.

Let us employ the G-S OP. Let $\mathbf{y}_1 := \mathbf{x}_1 = \begin{bmatrix}1 & 1 & 0 & 1\end{bmatrix}^\mathsf{T}$,

$$\begin{aligned}
\mathbf{y}_2 &:= \mathbf{x}_2 - P_{\mathbf{y}_1}(\mathbf{x}_2) = \begin{bmatrix}1 & -2 & 0 & 0\end{bmatrix}^\mathsf{T} + \frac{1}{3}\begin{bmatrix}1 & 1 & 0 & 1\end{bmatrix}^\mathsf{T} \\
&= \frac{1}{3}\begin{bmatrix}4 & -5 & 0 & 1\end{bmatrix}^\mathsf{T} \text{ and} \\
\mathbf{y}_3 &:= \mathbf{x}_3 - P_{\mathbf{y}_1}(\mathbf{x}_3) - P_{\mathbf{y}_2}(\mathbf{x}_3) \\
&= \begin{bmatrix}1 & 0 & -1 & 2\end{bmatrix}^\mathsf{T} - \frac{3}{3}\begin{bmatrix}1 & 1 & 0 & 1\end{bmatrix}^\mathsf{T} - \frac{1}{7}\begin{bmatrix}4 & -5 & 0 & 1\end{bmatrix}^\mathsf{T} \\
&= \frac{1}{7}\begin{bmatrix}-4 & -2 & -7 & 6\end{bmatrix}^\mathsf{T}.
\end{aligned}$$

Note that the subset $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ must be linearly independent since $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{y}_3$ are nonzero.

Further, let

$$\mathbf{u}_1 := \mathbf{y}_1/\sqrt{3},\ \mathbf{u}_2 := \sqrt{3}\,\mathbf{y}_2/\sqrt{14} \ \text{ and } \ \mathbf{u}_3 := \sqrt{7}\,\mathbf{y}_3/\sqrt{15}.$$

Then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis for the subspace W.

To extend $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ to an orthonormal basis for $V := \mathbb{K}^{4\times 1}$, we look for $\mathbf{y}_4 := \begin{bmatrix}\alpha_1 & \alpha_2 & \alpha_3 & \alpha_4\end{bmatrix}^\mathsf{T}$ which is orthogonal to the set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$, that is,

$$\alpha_1 + \alpha_2 + \alpha_4 = 0,\ \alpha_1 - 2\alpha_2 = 0 \ \text{ and } \alpha_1 - \alpha_3 + 2\alpha_4 = 0.$$

Letting $\alpha_1 := 2$, we obtain $\mathbf{y}_4 := \begin{bmatrix}2 & 1 & -4 & -3\end{bmatrix}^\mathsf{T}$. Then $\mathbf{y}_4$ is orthogonal to $\operatorname{span}\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\} = \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ as well.

Now let $\mathbf{u}_4 := \mathbf{y}_4/\|\mathbf{y}_4\| = \mathbf{y}_4/\sqrt{30}$.

Then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is an orthonormal basis for $\mathbb{K}^{4\times 1}$ which extends the orthonormal subset $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ of $\mathbb{K}^{4\times 1}$.

We point out an advantage of working with an orthonormal basis. Suppose $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a basis of $\mathbb{K}^{n\times 1}$. Then for $\mathbf{b} \in \mathbb{K}^{n\times 1}$, there are unique $\alpha_1, \ldots, \alpha_n$ such that $\mathbf{b} = \alpha_1\mathbf{x}_1 + \cdots + \alpha_n\mathbf{x}_n$. Finding these coefficients $\alpha_1, \ldots, \alpha_n$ is not always easy. In fact, $\begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix}^\mathsf{T}$ is the unique column vector satisfying the linear system

$$\mathbf{A}\begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix}^\mathsf{T} = \mathbf{b}, \text{ where } \mathbf{A} := \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix},$$

and we would have to find this vector either by the GEM or by the Cramer Rule (involving $n+1$ determinants of size n).

On the other hand, suppose $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal basis of $\mathbb{K}^{n\times 1}$. If $\mathbf{b} \in \mathbb{K}^{n\times 1}$ and $\mathbf{b} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n$, then $\alpha_j = \langle \mathbf{u}_j, \mathbf{b}\rangle$ for $j = 1, \ldots, n$ by the orthonormality, so that

$$\mathbf{b} = \langle \mathbf{u}_1, \mathbf{b}\rangle\mathbf{u}_1 + \cdots + \langle \mathbf{u}_n, \mathbf{b}\rangle\mathbf{u}_n.$$

For instance, consider the ordered orthonormal basis $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4)$ of $\mathbb{K}^{4\times 1}$ which we have just constructed, where

$$\mathbf{u}_1 := \begin{bmatrix}1 & 1 & 0 & 1\end{bmatrix}^\mathsf{T}/\sqrt{3},$$

$$\mathbf{u}_2 := \begin{bmatrix}4 & -5 & 0 & 1\end{bmatrix}^\mathsf{T}/\sqrt{42},$$

$$\mathbf{u}_3 := \begin{bmatrix}-4 & -2 & -7 & 6\end{bmatrix}^\mathsf{T}/\sqrt{105},$$

$$\mathbf{u}_4 := \begin{bmatrix}2 & 1 & -4 & -3\end{bmatrix}^\mathsf{T}/\sqrt{30}.$$

Let $\mathbf{b} := \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^\mathsf{T} \in \mathbb{K}^{4\times 1}$.

Then there are unique $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathbb{K}$ such that $\mathbf{b} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \alpha_4\mathbf{u}_4$. In fact,

$$\alpha_1 = \langle \mathbf{u}_1, \mathbf{b}\rangle = 3/\sqrt{3} = \sqrt{3},$$

$$\alpha_2 = \langle \mathbf{u}_2, \mathbf{b}\rangle = 0,$$

$$\alpha_3 = \langle \mathbf{u}_3, \mathbf{b}\rangle = -7/\sqrt{105} = -\sqrt{7}/\sqrt{15},$$

$$\alpha_4 = \langle \mathbf{u}_4, \mathbf{b}\rangle = -4/\sqrt{30}.$$

Analogue for Row Vectors

We have defined an inner product as a function from $\mathbb{K}^{n\times 1} \times \mathbb{K}^{n\times 1}$ to $\mathbb{K}$. We can also define a similar function from $\mathbb{K}^{1\times n} \times \mathbb{K}^{1\times n}$ to $\mathbb{K}$ as follows.

Consider row vectors $\mathbf{x} := \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$, $\mathbf{y} := \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}$ in $\mathbb{K}^{1\times n}$. The **inner product** of $\mathbf{x}$ with $\mathbf{y}$ is defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle := \overline{\mathbf{x}}\mathbf{y}^{\mathsf{T}} = \overline{x}_1 y_1 + \cdots + \overline{x}_n y_n.$$

Also, we can introduce the concepts of orthogonality and orthonormality of row vectors, and obtain a Gram-Schmidt Orthogonalization Process for row vectors.

After a detour of inner products and orthonormal sets, we come back to the matrix eigenvalue problem. We shall show that if the scalars are complex numbers, then every square matrix $\mathbf{A}$ can be 'upper triangularized', that is, it is similar to an upper triangular matrix $\mathbf{B} := \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. In fact, we shall show that the invertible matrix $\mathbf{P}$ can be chosen to be of a particularly nice type. We now describe what we mean by nice.

A matrix $\mathbf{U} \in \mathbb{K}^{n\times n}$ is called **unitary** if the columns of $\mathbf{U}$ form an orthonormal subset of $\mathbb{K}^{n\times 1}$. In that case, the columns are in fact an orthonormal basis for $\mathbb{K}^{n\times 1}$.

Proposition

A matrix is unitary if and only if it is invertible and its inverse is the same as its adjoint.

Proof. Let $\mathbf{U}$ be unitary. Then $\operatorname{rank} \mathbf{U} = n$ since the n columns $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of $\mathbf{U}$ form a basis for $\mathbb{K}^{n\times 1}$. Hence $\mathbf{U}$ is invertible. Further, because of the orthonormality,

$$\mathbf{U}^*\mathbf{U} = \begin{bmatrix} \mathbf{u}_1^* \\ \vdots \\ \mathbf{u}_n^* \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^*\mathbf{u}_1 & \cdots & \mathbf{u}_1^*\mathbf{u}_n \\ \vdots & \vdots & \vdots \\ \mathbf{u}_n^*\mathbf{u}_1 & \cdots & \mathbf{u}_n^*\mathbf{u}_n \end{bmatrix} = \mathbf{I}.$$

It follows that $\mathbf{U}\mathbf{U}^* = \mathbf{I}$ as well. Hence $\mathbf{U}^{-1} = \mathbf{U}^*$.

Conversely, the above calculation shows that if a square matrix $\mathbf{U}$ satisfies $\mathbf{U}^*\mathbf{U} = \mathbf{I}$, then its columns form an orthonormal set, that is, $\mathbf{U}$ is a unitary matrix. $\square$

We note that if $\mathbf{A}$ and $\mathbf{B}$ are unitary, then so is $\mathbf{AB}$ since $(\mathbf{AB})^*(\mathbf{AB}) = (\mathbf{B}^*\mathbf{A}^*)(\mathbf{AB}) = \mathbf{B}(\mathbf{A}^*\mathbf{A})\mathbf{B} = \mathbf{B}^*\mathbf{B} = \mathbf{I}$.

Examples

(i) The $n \times n$ identity matrix $\mathbf{I}_n := \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix}$ is unitary.

(ii) For $\theta \in \mathbb{R}$, the matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$ is unitary.

It represents a rotation about the x_1-axis in $\mathbb{R}^{3\times 1}$.

(iii) The matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ is unitary. It represents reflection along the x_3-axis in $\mathbb{R}^{3\times 1}$.

(iv) If $\mathbb{K} = \mathbb{C}$, then the matrix $\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -i/\sqrt{2} & i/\sqrt{2} & 0 \\ 0 & 0 & i \end{bmatrix}$ is unitary.

(v) The matrix $\begin{bmatrix} 1/\sqrt{3} & 4/\sqrt{42} & -4/\sqrt{105} & 2/\sqrt{30} \\ 1/\sqrt{3} & -5/\sqrt{42} & -2/\sqrt{105} & 1/\sqrt{30} \\ 0 & 0 & -7/\sqrt{105} & -4/\sqrt{30} \\ 1/\sqrt{3} & 1/\sqrt{42} & 6/\sqrt{105} & -3/\sqrt{30} \end{bmatrix}$,

whose columns are obtained by orthonormalizing the column vectors $\begin{bmatrix} 1 & 1 & 0 & 1 \end{bmatrix}^{\mathsf{T}}, \begin{bmatrix} 1 & -2 & 0 & 0 \end{bmatrix}^{\mathsf{T}}, \begin{bmatrix} 1 & 0 & -1 & 2 \end{bmatrix}^{\mathsf{T}}$ and $\begin{bmatrix} 2 & 1 & -4 & -3 \end{bmatrix}^{\mathsf{T}}$ in $\mathbb{K}^{4\times 1}$ is unitary.

(vi) We shall consider an interesting unitary matrix later.

Let $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{n\times n}$. We say that $\mathbf{A}$ is **unitarily similar** to $\mathbf{B}$ if there is a unitary matrix $\mathbf{U}$ such that $\mathbf{B} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$. Also, $\mathbf{A}$ is called **unitarily diagonalizable** if it is unitarily similar to a diagonal matrix.

We proved in Lecture 10 that a matrix $\mathbf{A}$ is similar to a matrix $\mathbf{B}$ if and only if there is an ordered basis E for $\mathbb{K}^{n\times 1}$ such that $\mathbf{B}$ is the matrix of the linear map $T_{\mathbf{A}} : \mathbb{K}^{n\times 1} \to \mathbb{K}^{n\times 1}$ with respect to E. We prove an analogous result below.

Proposition

Let $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{n\times n}$. Then $\mathbf{A}$ is unitarily similar to $\mathbf{B}$ if and only if there is an ordered orthonormal basis E for $\mathbb{K}^{n\times 1}$ such that $\mathbf{B}$ is the matrix of the linear transformation $T_{\mathbf{A}} : \mathbb{K}^{n\times 1} \to \mathbb{K}^{n\times 1}$ with respect to E.
In fact, $\mathbf{B} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$ if and only if the columns of $\mathbf{U}$ form an ordered orthonormal basis, say E, for $\mathbb{K}^{n\times 1}$ and $\mathbf{B} = \mathbf{M}_E^E(T_{\mathbf{A}})$.

Proof. Let $\mathbf{B} := [b_{jk}]$. Now $\mathbf{A}$ is unitarily similar to $\mathbf{B}$ if and only if there is a unitary matrix $\mathbf{U}$ such that $\mathbf{AU} = \mathbf{UB}$. This is the case if and only if there is an ordered orthonormal basis $E := (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ for $\mathbb{K}^{n\times 1}$ such that

$$\mathbf{A}\begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{bmatrix}\begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \vdots & \vdots \\ b_{n1} & \cdots & b_{nn}\end{bmatrix}.$$

The kth column of LHS is $\mathbf{Au}_k$ and the kth column of RHS is the linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_n$ with coefficients from the k column of $\mathbf{B}$. Thus $\mathbf{Au}_k = b_{1k}\mathbf{u}_1 + \cdots + b_{nk}\mathbf{u}_n$ for $k = 1, \ldots, n$. This means the kth column of $\mathbf{M}_E^E(T_\mathbf{A})$ is the kth column $\begin{bmatrix} b_{1k} & \cdots & b_{nk}\end{bmatrix}^\mathsf{T}$ of $\mathbf{B}$, $k = 1, \ldots, n$, that is, $\mathbf{B} = \mathbf{M}_E^E(T_\mathbf{A})$. □

The above result says that just as $\mathbf{A}$ is the matrix of the linear transformation $T_\mathbf{A}$ defined by $\mathbf{A}$ with respect to the standard ordered basis $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ for $\mathbb{K}^{n\times 1}$, the matrix $\mathbf{B} := \mathbf{U}^{-1}\mathbf{AU}$ is the matrix of the same linear transformation $T_\mathbf{A}$ with respect to the ordered orthonormal basis for $\mathbb{K}^{n\times 1}$ consisting of the columns of $\mathbf{U}$.

In Lecture 10, we saw that a matrix $\mathbf{A} \in \mathbb{K}^{n \times n}$ is diagonalizable (that is, $\mathbf{A}$ is similar to a diagonal matrix) if and only if there is a basis for $\mathbb{K}^{n \times 1}$ consisting of eigenvectors of $\mathbf{A}$. Similarly, we can prove the following result.

Proposition

A matrix $\mathbf{A} \in \mathbb{K}^{n \times n}$ is unitarily diagonalizable if and only if there is an orthonormal basis for $\mathbb{K}^{n \times 1}$ consisting of eigenvectors of $\mathbf{A}$.

In fact, $\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{D}$, where $\mathbf{U} := \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$ and $\mathbf{D} := \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \iff \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{K}^{n \times 1}$ and $\mathbf{A}\mathbf{u}_k = \lambda_k \mathbf{u}_k$ for $k = 1, \ldots, n$.

If $\mathbb{K} := \mathbb{C}$, then we can nicely characterise matrices that can be unitarily diagonalized. As a preparation, we prove a powerful result which says that every $\mathbf{A} \in \mathbb{C}^{n \times n}$ can be unitarily 'upper triangularized'.

Theorem (Schur)

Let $\mathbf{A} \in \mathbb{C}^{n\times n}$. Then $\mathbf{A}$ is unitarily similar to an upper triangular matrix.

Proof. We use induction on n. If $n = 1$, then there is nothing to prove. Let $n \geq 2$ and assume that the result holds for all $(n-1)\times(n-1)$ matrices with complex scalars.

We shall show that there is an $n\times n$ unitary matrix $\mathbf{U}$ and also an $n\times n$ upper triangular matrix $\mathbf{B}$ such that $\mathbf{A} = \mathbf{U}\mathbf{B}\mathbf{U}^*$.

By the Fundamental Theorem of Algebra, the characteristic polynomial of $\mathbf{A}$ has a root in $\mathbb{C}$. Hence there is $\lambda_1 \in \mathbb{C}$ and there is nonzero $\mathbf{x}_1 \in \mathbb{C}^{n\times 1}$ such that $\mathbf{A}\,\mathbf{x}_1 = \lambda_1\mathbf{x}_1$. We may assume WLOG that $\|\mathbf{x}_1\| = 1$.

We extend the orthonormal set $\{\mathbf{x}_1\}$ in $\mathbb{C}^{n\times 1}$ to an ordered orthonormal basis $E := (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ for $\mathbb{C}^{n\times 1}$.

Define $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix}$. Then $\mathbf{X}$ is a unitary matrix, so that $\mathbf{X}^{-1} = \mathbf{X}^*$.

Consider the linear transformation $T_{\mathbf{A}} : \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$ defined by $T_{\mathbf{A}}(\mathbf{x}) := \mathbf{A}\,\mathbf{x}$. Let $\mathbf{C}$ denote the matrix $\mathbf{M}_E^E(T)$ of this linear map with respect to the basis E. Then $\mathbf{A}\,\mathbf{P} = \mathbf{P}\,\mathbf{C}$, that is, $\mathbf{A} = \mathbf{P}\,\mathbf{C}\,\mathbf{P}^*$. Also, since $\mathbf{A}\,\mathbf{x}_1 = \lambda_1\mathbf{x}_1$, we obtain

$$\mathbf{C} = \left[\begin{array}{c|ccc} \lambda_1 & \alpha_2 & \cdots & \alpha_n \\ \hline 0 & & & \\ \vdots & & \mathbf{A}_1 & \\ 0 & & & \end{array}\right],$$

where $\alpha_2, \ldots, \alpha_n \in \mathbb{C}$ and $\mathbf{A}_1 \in \mathbb{C}^{(n-1)\times(n-1)}$. By the induction hypothesis, $\mathbf{A}_1 = \mathbf{P}_1\mathbf{B}_1\mathbf{P}_1^*$, where $\mathbf{P}_1$ is an $(n-1)\times(n-1)$ unitary matrix and $\mathbf{B}_1 = \mathbf{P}_1^*\mathbf{A}_1\mathbf{P}_1$ is an $(n-1)\times(n-1)$ upper triangular matrix.
We now ‘border’ the unitary matrix $\mathbf{P}_1$ as follows.

Define
$$\mathbf{U}_1 := \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1 & \\ 0 & & & \end{array}\right].$$

Clearly, $\mathbf{U}_1$ is unitary. Now define $\mathbf{B} := \mathbf{U}_1^*\mathbf{C}\mathbf{U}_1$. Then

$$\mathbf{B} = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1 & \\ 0 & & & \end{array}\right]^* \left[\begin{array}{c|ccc} \lambda_1 & \alpha_2 & \cdots & \alpha_n \\ \hline 0 & & & \\ \vdots & & \mathbf{A}_1 & \\ 0 & & & \end{array}\right] \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1 & \\ 0 & & & \end{array}\right]$$

that is,

$$\left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1^* & \\ 0 & & & \end{array}\right] \left[\begin{array}{c|ccc} \lambda_1 & \beta_2 & \cdots & \beta_n \\ \hline 0 & & & \\ \vdots & & \mathbf{A}_1\mathbf{P}_1 & \\ 0 & & & \end{array}\right] = \left[\begin{array}{c|ccc} \lambda_1 & \beta_2 & \cdots & \beta_n \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1^*\mathbf{A}_1\mathbf{P}_1 & \\ 0 & & & \end{array}\right],$$

where $\beta_2, \ldots, \beta_n \in \mathbb{C}$. Now the matrix

$$\mathbf{B} = \mathbf{U}_1^* \mathbf{C} \mathbf{U}_1 = \left[\begin{array}{c|ccc} \lambda_1 & \beta_2 & \cdots & \beta_n \\ \hline 0 & & & \\ \vdots & & \mathbf{P}_1^* \mathbf{A}_1 \mathbf{P}_1 & \\ 0 & & & \end{array}\right] = \left[\begin{array}{c|ccc} \lambda_1 & \beta_2 & \cdots & \beta_n \\ \hline 0 & & & \\ \vdots & & \mathbf{B}_1 & \\ 0 & & & \end{array}\right]$$

is upper triangular since $\mathbf{B}_1$ is upper triangular. Also, since $\mathbf{C} = \mathbf{U}_1 \mathbf{B}\, \mathbf{U}_1^*$, we see that

$$\mathbf{A} = \mathbf{P}\,\mathbf{C}\,\mathbf{P}^* = \mathbf{P}(\mathbf{U}_1 \mathbf{B}\, \mathbf{U}_1^*)\mathbf{P}^* = (\mathbf{P}\mathbf{U}_1)\mathbf{B}(\mathbf{P}\,\mathbf{U}_1)^*.$$

Because $\mathbf{P}$ and $\mathbf{U}_1$ are unitary, so is $\mathbf{U} := \mathbf{P}\mathbf{U}_1$, and we obtain $\mathbf{A} = \mathbf{U}\mathbf{B}\mathbf{U}^*$, as desired. $\square$