Math 504: Advanced Linear Algebra

Hugo Woerdeman, with edits by Darij Grinberg*

September 24, 2021 (unfinished!)

Contents

Math 504 Lecture 2

1. Unitary matrices ([HorJoh13, §2.1]) (cont'd)

Recall:

- An $n \times k$-matrix A is said to be an **isometry** if $A^* A = I_k$. (The notation I_k means the $k \times k$ identity matrix.)
- An $n \times k$-matrix A is an isometry if and only if its columns form an orthonormal tuple of vectors.

1.1. Unitary matrices

Definition 1.1.1. A matrix $U \in \mathbb{C}^{n \times k}$ is said to be **unitary** if both U and U^* are isometries.

*Drexel University, Korman Center, 15 S 33rd Street, Philadelphia PA, 19104, USA

Example 1.1.2. **(a)** The matrix $A = \dfrac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is unitary. Indeed, it is easy to see that A is an isometry, but A^* is therefore also an isometry, since $A^* = A$. Thus, A is unitary.

(b) A 1×1-matrix $\begin{pmatrix} \lambda \end{pmatrix}$ is unitary if and only if $|\lambda| = 1$.

(c) For any $n \in \mathbb{N}$, the identity matrix I_n is an isometry and thus unitary.

(d) Let $n \in \mathbb{N}$. Let σ be a permutation of $[n] = \{1, 2, \ldots, n\}$. That is, σ is a bijective map from $[n]$ to $[n]$.

Let P_σ be the **permutation matrix** of σ; this is the $n \times n$-matrix whose $(\sigma(j), j)$-th entry is 1 for all $j \in [n]$, and whose all other entries are 0.

For instance, if $n = 3$ and $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$, then

$$P_\sigma = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

The permutation matrix P_σ is always unitary (for any n and any σ). Indeed, the inverse of P_σ is $P_{\sigma^{-1}}$, but the conjugate transpose of P_σ is also $P_{\sigma^{-1}}$. Thus, $P_\sigma^* = P_\sigma^{-1}$, so $P_\sigma^* P_\sigma = I_n$. This shows that P_σ is an isometry. Similarly, $P_\sigma P_\sigma^* = I_n$, so that P_σ^* is an isometry, and thus P_σ is unitary.

(e) A diagonal matrix $\operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$ is unitary if and only if

$$|\lambda_1| = |\lambda_2| = \cdots = |\lambda_n| = 1.$$

Indeed,

$$\begin{aligned}
&\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}^* \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \\
&= \begin{pmatrix} \overline{\lambda_1} & 0 & \cdots & 0 \\ 0 & \overline{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \overline{\lambda_n} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \\
&= \begin{pmatrix} \overline{\lambda_1}\lambda_1 & 0 & \cdots & 0 \\ 0 & \overline{\lambda_2}\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \overline{\lambda_n}\lambda_n \end{pmatrix} = \begin{pmatrix} |\lambda_1|^2 & 0 & \cdots & 0 \\ 0 & |\lambda_2|^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & |\lambda_n|^2 \end{pmatrix}.
\end{aligned}$$

Theorem 1.1.3. Let $U \in \mathbb{C}^{n \times k}$ be a matrix. The following six statements are equivalent:

- $\mathcal{A}$: The matrix U is unitary.
- $\mathcal{B}$: The matrices U and U^* are isometries.
- $\mathcal{C}$: We have $UU^* = I_n$ and $U^*U = I_k$.
- $\mathcal{D}$: The matrix U is square (that is, $n = k$) and invertible and satisfies $U^{-1} = U^*$.
- $\mathcal{E}$: The columns of U form an orthonormal basis of $\mathbb{C}^n$.
- $\mathcal{F}$: The matrix U is square (that is, $n = k$) and is an isometry.

Proof. $\mathcal{A} \Longleftrightarrow \mathcal{B}$ follows from the def of "unitary".
$\mathcal{B} \Longleftrightarrow \mathcal{C}$ follows from the def of "isometry", since $(U^*)^* = U$.
$\mathcal{D} \Longrightarrow \mathcal{C}$ is obvious.
$\mathcal{C} \Longrightarrow \mathcal{D}$ follows from the fact that any invertible matrix is square.
The other implications are a bit harder:

- $\mathcal{D} \Longrightarrow \mathcal{E}$: Assume that $\mathcal{D}$ holds. Thus, U is an isometry (since $U^*U = I_k$). So, from a result from last time, we see that the columns of U form an orthonormal tuple of vectors. But they also form a basis of $\mathbb{C}^n$, because U is invertible. So $\mathcal{E}$ holds.
- $\mathcal{E} \Longrightarrow \mathcal{D}$: Assume that $\mathcal{E}$ holds. Then, the columns of U form an orthonormal basis, hence an orthonormal tuple. Thus, U is an isometry (by what we know from last lecture). Furthermore, the columns of U form a basis, so there are precisely n of them (since any basis of $\mathbb{C}^n$ has n vectors). This shows that U is square. Finally, U is invertible, since its columns form a basis. So $\mathcal{D}$ holds.
- $\mathcal{D} \Longrightarrow \mathcal{F}$: Easy.
- $\mathcal{F} \Longrightarrow \mathcal{D}$: It is known that a square matrix A that is left invertible (i.e., there is a matrix B such that $BA = I$) is always invertible. Since U is an isometry, we have $U^*U = I_k$, so that U is left-invertible. Thus, U is invertible, and its inverse is $U^{-1} = U^*$. So $\mathcal{D}$ holds.

This proves $\mathcal{A} \Longleftrightarrow \mathcal{B} \Longleftrightarrow \mathcal{C} \Longleftrightarrow \mathcal{D} \Longleftrightarrow \mathcal{E} \Longleftrightarrow \mathcal{F}$. □

This theorem shows that any unitary matrix is square. In contrast, an isometry can be a tall matrix (but not a wide matrix).

The set of all unitary $n \times n$-matrices is called the **n-th unitary group**, and is denoted $\mathrm{U}_n(\mathbb{C})$. It is a group under multiplication.

If U is a unitary matrix, then $|\det U| = 1$ and any eigenvalue λ of U has $|\lambda| = 1$.

1.1.1. Block matrices

Definition 1.1.4. Let $A \in \mathbb{C}^{n\times p}$, $B \in \mathbb{C}^{n\times q}$, $C \in \mathbb{C}^{m\times p}$, $D \in \mathbb{C}^{m\times q}$ be four matrices. Then, $\left(\begin{array}{cc} A & B \\ C & D \end{array}\right)$ means the matrix

$$\left(\begin{array}{cccccc} A_{1,1} & A_{1,2} & \cdots & B_{1,1} & B_{1,2} & \cdots \\ A_{2,1} & A_{2,2} & \cdots & B_{2,1} & B_{2,2} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ C_{1,1} & C_{1,2} & \cdots & D_{1,1} & D_{1,2} & \cdots \\ C_{2,1} & C_{2,2} & \cdots & D_{2,1} & D_{2,2} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{array}\right) \in \mathbb{C}^{(n+m)\times(p+q)}$$

(where $M_{i,j}$ means the (i,j)-th entry of a matrix M).

This matrix is called the **block matrix** formed of A, B, C and D.

Similar notations will be used to glue together more than 4 matrices.

This block matrix notation is more than just a convenient notation. Indeed, multiplication of matrices plays along with it:

Proposition 1.1.5. For any matrices $A, B, C, D, A', B', C', D'$, we have

$$\left(\begin{array}{cc} A & B \\ C & D \end{array}\right)\left(\begin{array}{cc} A' & B' \\ C' & D' \end{array}\right) = \left(\begin{array}{cc} AA'+BC' & AB'+BD' \\ CA'+DC' & CB'+DD' \end{array}\right),$$

provided that the block matrices and the products make sense.

Again, a similar rule holds for larger block matrices.

A particularly well-behaved family of block matrices are the **block-diagonal matrices**. These are the block matrices of the form

$$\left(\begin{array}{cccc} A(1,1) & 0 & \cdots & 0 \\ 0 & A(2,2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A(n,n) \end{array}\right),$$

where each $A(i,i)$ is a square matrix, and where the 0's mean zero matrices of appropriate dimension.

These block-diagonal matrices can be muliplied "diagonal block by diagonal block":

$$\begin{pmatrix} A(1,1) & 0 & \cdots & 0 \\ 0 & A(2,2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A(n,n) \end{pmatrix} \begin{pmatrix} B(1,1) & 0 & \cdots & 0 \\ 0 & B(2,2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B(n,n) \end{pmatrix}$$
$$= \begin{pmatrix} A(1,1)B(1,1) & 0 & \cdots & 0 \\ 0 & A(2,2)B(2,2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A(n,n)B(n,n) \end{pmatrix},$$

provided that

$$\text{size of } A(i,i) = \text{size of } B(i,i) \text{ for each } i.$$

Example: $\begin{pmatrix} a & 0 & 0 \\ 0 & b & c \\ 0 & d & e \end{pmatrix} \begin{pmatrix} a' & 0 & 0 \\ 0 & b' & c' \\ 0 & d' & e' \end{pmatrix} = \begin{pmatrix} aa' & 0 & 0 \\ 0 & cd'+bb' & ce'+bc' \\ 0 & d'e+db' & e'e+dc' \end{pmatrix}$

Non-example: $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 4 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 \\ 8 & 6 & 3 \\ 4 & 3 & 4 \end{pmatrix}$

Proposition 1.1.6. Let $A_1, A_2, \ldots, A_u$ be a bunch of square matrices. Then, the block-diagonal matrix $\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_u \end{pmatrix}$ is unitary if and only $A_1, A_2, \ldots, A_u$ are unitary.

1.2. The Gram–Schmidt process

Theorem 1.2.1 (Gram–Schmidt process)**.** Let $(v_1, v_2, \ldots, v_m)$ be a linearly independent tuple of vectors in $\mathbb{C}^n$.

Then, there is an orthogonal tuple $(z_1, z_2, \ldots, z_m)$ of vectors in $\mathbb{C}^n$ that satisfies

$$\operatorname{span}\{v_1, v_2, \ldots, v_j\} = \operatorname{span}\{z_1, z_2, \ldots, z_j\} \qquad \text{for all } j \in [m].$$

Furthermore, such a tuple $(z_1, z_2, \ldots, z_m)$ can be constructed by the following recursive process:

- For each $p \in [m]$, if the first $p-1$ entries $z_1, z_2, \ldots, z_{p-1}$ of this tuple have already been constructed, then we define the p-th entry z_p by

$$z_p = v_p - \sum_{k=1}^{p-1} \frac{\langle v_p, z_k \rangle}{\langle z_k, z_k \rangle} z_k.$$

(Note that when $p = 1$, the sum on the RHS is an empty sum, so this equality simply becomes $z_1 = v_1$.)

Roughly speaking, the claim of this theorem is that if we start with any linearly independent tuple $(v_1, v_2, \ldots, v_m)$ of vectors in $\mathbb{C}^n$, then we can make this tuple orthogonal by tweaking it as follows:

- we leave v_1 unchanged;
- we modify v_2 by subtracting an appropriate scalar multiple of v_1;
- we modify v_3 by subtracting an appropriate linear combination of v_1 and v_2;
- and so on.

The above recursive formula tells us which scalar multiples / linear combinations to take.

Example 1.2.2.

$$\begin{aligned}
z_1 &= v_1; \\
z_2 &= v_2 - \frac{\langle v_2, z_1 \rangle}{\langle z_1, z_1 \rangle} z_1; \\
z_3 &= v_3 - \frac{\langle v_3, z_1 \rangle}{\langle z_1, z_1 \rangle} z_1 - \frac{\langle v_3, z_2 \rangle}{\langle z_2, z_2 \rangle} z_2; \\
&\ldots
\end{aligned}$$