Math 504: Advanced Linear Algebra

Hugo Woerdeman, with edits by Darij Grinberg*

September 24, 2021 (unfinished!)

Contents

Math 504 Lecture 3

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

1.1. Administrativa

Next topics after today:

- Normal matrices and the spectral thm for them (lec 4) <– Scribe: Phil (on Monday)
- The Cayley–Hamilton theorem (lec 5) <– Scribe: Atharv (on Wednesday)
- Sylvester's equation $AX - XB = C$ (lec 5–6) <– Scribe: Hunter (on Friday)

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

- Jordan canonical form (lec 6–7) <– Scribe: Juliane

Who scribes?

1.2. The Gram–Schmidt process (cont'd)

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.

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

See the notes for an actual example (with numbers).

Proof. We must show three things:

1. The $z_1, z_2, \ldots, z_m$ constructed by the recursive process exist.
2. They satisfy span $\{v_1, v_2, \ldots, v_j\} =$ span $\{z_1, z_2, \ldots, z_j\}$ for all $j \in [m]$.
3. The tuple $(z_1, z_2, \ldots, z_m)$ is orthogonal.

We will first prove statements 1 and 2 in lockstep:

Claim 1: For each $p \in \{0, 1, \ldots, m\}$, the vectors $z_1, z_2, \ldots, z_p$ are well-defined and satisfy

$$\operatorname{span}\{v_1, v_2, \ldots, v_p\} = \operatorname{span}\{z_1, z_2, \ldots, z_p\}.$$

[*Proof of Claim 1:* We induct on p. The *base case* ($p = 0$) is obvious.

For the *induction step*, we fix $p \in [m]$, and we assume that the claim holds for $p - 1$. In other words, we assume that the vectors $z_1, z_2, \ldots, z_{p-1}$ are well-defined and satisfy

$$\operatorname{span}\{v_1, v_2, \ldots, v_{p-1}\} = \operatorname{span}\{z_1, z_2, \ldots, z_{p-1}\}. \tag{1}$$

We must show that the vectors $z_1, z_2, \ldots, z_p$ are well-defined and satisfy

$$\operatorname{span}\{v_1, v_2, \ldots, v_p\} = \operatorname{span}\{z_1, z_2, \ldots, z_p\}. \tag{2}$$

First, we use (1) to conclude that $(z_1, z_2, \ldots, z_{p-1})$ is linearly independent, since their span has dimension $p - 1$. Thus, in particular, for each $k \in [p-1]$, we have $z_k \neq 0$, so that $\langle z_k, z_k \rangle > 0$. Thus, in the equality

$$z_p = v_p - \sum_{k=1}^{p-1} \frac{\langle v_p, z_k \rangle}{\langle z_k, z_k \rangle} z_k, \tag{3}$$

the denominators are nonzero, so that z_p is well-defined. Thus, $z_1, z_2, \ldots, z_p$ are well-defined.

Now, (2) follows from (1) and (3). (See the notes for details.) So the induction is complete, and Claim 1 is proven.]

Now it remains to prove statement 3:

Claim 2: For any $j \in \{0, 1, \ldots, m\}$, the tuple $(z_1, z_2, \ldots, z_j)$ is orthogonal.

[*Proof of Claim 2:* Induct on j. The *base case* is trivial.

Induction step: Let $p \in [m]$. Assume (as the IH) that Claim 2 holds for $j = p - 1$. We must show that it holds for $j = p$.

By our IH, the tuple $(z_1, z_2, \ldots, z_{p-1})$ is orthogonal. It thus remains to show that z_p is orthogonal to each z_a with $a \in [p-1]$. To that purpose, we fix some $a \in [p-1]$, and we shall show that $\langle z_p, z_a \rangle = 0$. To wit:

$$\begin{aligned}
\langle z_p, z_a \rangle &= \left\langle v_p - \sum_{k=1}^{p-1} \frac{\langle v_p, z_k \rangle}{\langle z_k, z_k \rangle} z_k, z_a \right\rangle \qquad \left(\text{since } z_p = v_p - \sum_{k=1}^{p-1} \frac{\langle v_p, z_k \rangle}{\langle z_k, z_k \rangle} z_k \right) \\
&= \langle v_p, z_a \rangle - \sum_{k=1}^{p-1} \frac{\langle v_p, z_k \rangle}{\langle z_k, z_k \rangle} \underbrace{\langle z_k, z_a \rangle}_{\substack{=0 \text{ unless } k=a \\ (\text{because the tuple } (z_1, z_2, \ldots, z_{p-1}) \\ \text{is orthogonal})}} \\
&= \langle v_p, z_a \rangle - \frac{\langle v_p, z_a \rangle}{\langle z_a, z_a \rangle} \langle z_a, z_a \rangle = 0,
\end{aligned}$$

so that $z_p \perp z_a$, as desired. This completes the induction.]

With Claim 1 and Claim 2 proved, the proof of the theorem is complete. □

What if the original tuple $(v_1, v_2, \ldots, v_m)$ is not linearly independent? We can adapt our above theorem to this case:

Theorem 1.2.3 (Gram–Schmidt process, take 2)**.** Let $(v_1, v_2, \ldots, v_m)$ be any tuple of vectors in $\mathbb{C}^n$ with $m \leq n$.

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

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

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 as follows:
  - If $v_p - \sum\limits_{k=1}^{p-1} \frac{\langle v_p, z_k \rangle}{\langle z_k, z_k \rangle} z_k \neq 0$, then we set

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

- If $v_p - \sum_{k=1}^{p-1} \frac{\langle v_p, z_k \rangle}{\langle z_k, z_k \rangle} z_k = 0$, then we pick an arbitrary nonzero vector b that is orthogonal to all of $z_1, z_2, \ldots, z_{p-1}$ (such a b exists, as previously shown), and we set
$$z_p := b.$$

Proof. This requires a variant of the proof of the previous theorem; see the notes. □

Corollary 1.2.4. Let $(v_1, v_2, \ldots, v_m)$ be any tuple of vectors in $\mathbb{C}^n$ with $m \leq n$.

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

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

Proof. Use the previous theorem to obtain an orthogonal tuple $(z_1, z_2, \ldots, z_m)$ with this property. Then, replace it by $\left(\frac{1}{||z_1||} z_1, \ \frac{1}{||z_2||} z_2, \ \ldots, \ \frac{1}{||z_m||} z_m \right)$. □

1.3. QR factorization

Recall that an isometry is a matrix whose columns form an orthonormal tuple.

Theorem 1.3.1 (QR factorization, isometry version)**.** Let $A \in \mathbb{C}^{n \times m}$ satisfy $n \geq m$. Then, there exists an isometry $Q \in \mathbb{C}^{n \times m}$ and an upper-triangular matrix $R \in \mathbb{C}^{m \times m}$ such that $A = QR$.

The pair (Q, R) in this theorem is called a **QR factorization** of A.

Example 1.3.2. Let

$$A = \begin{pmatrix} 1 & 0 & 1 & 2 \\ 1 & -2 & 0 & 2 \\ 1 & 0 & 1 & 0 \\ 1 & -2 & 0 & 0 \end{pmatrix} \in \mathbb{C}^{4 \times 4}.$$

Then, one QR factorization of A is

$$A = \underbrace{\begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \end{pmatrix}}_{=Q} \underbrace{\begin{pmatrix} 2 & -2 & 1 & 0 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}}_{=R}.$$

There are others.

Proof of Theorem. Let $A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m}$ denote the m columns of A.

Applying the previous corollary to $(v_1, v_2, \ldots, v_m) = (A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,m})$, we conclude that there exists an orthonormal tuple $(q_1, q_2, \ldots, q_m)$ of vectors in $\mathbb{C}^n$ such that

$$\operatorname{span}\{A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,j}\} \subseteq \operatorname{span}\{q_1, q_2, \ldots, q_j\} \qquad \text{for all } j \in [m].$$

Consider this tuple. Let $Q \in \mathbb{C}^{n \times m}$ be the matrix whose columns are $q_1, q_2, \ldots, q_m$. Then, Q is an isometry.

Now, I claim that $A = QR$ for some upper-triangular R. Indeed, for each $j \in [m]$, we have

$$A_{\bullet,j} \in \operatorname{span}\{A_{\bullet,1}, A_{\bullet,2}, \ldots, A_{\bullet,j}\} \subseteq \operatorname{span}\{q_1, q_2, \ldots, q_j\},$$

so that

$$A_{\bullet,j} = r_{1,j}q_1 + r_{2,j}q_2 + \cdots + r_{j,j}q_j \qquad \text{for some } r_{1,j}, r_{2,j}, \ldots, r_{j,j} \in \mathbb{C}.$$

This shows that $A = QR$ for the upper-triangular matrix

$$R = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,m} \\ 0 & r_{2,2} & \cdots & r_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{m,m} \end{pmatrix}.$$

□

2. Schur triangularization

2.1. Similarity of matrices

Definition 2.1.1. Let $\mathbb{F}$ be a field. Let A and B be two $n \times n$-matrices over $\mathbb{F}$. We say that A is **similar** to B if there exists an invertible matrix $W \in \mathbb{F}^{n \times n}$ such that $B = WAW^{-1}$.

Example 2.1.2. The matrix $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is similar to the matrix $\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, since

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} = W \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} W^{-1} \qquad \text{for } W = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Remark 2.1.3. If you think of matrices as representing linear maps, then similarity has a much more fundamental meaning: A matrix is similar to another if and only if the two matrices represent the same endomorphism of $\mathbb{F}^n$ (i.e., linear map $\mathbb{F}^n \to \mathbb{F}^n$) with respect to two bases.

The relation "similar" is an equivalence relation: i.e., it is

- **reflexive:** Any matrix $A \in \mathbb{F}^{n\times n}$ is similar to itself.
- **symmetric:** If A is similar to B, then B is similar to A.
- **transitive:** If A is similar to B, and B is similar to C, then A is similar to C.

To check these, recall that $(UV)^{-1} = V^{-1}U^{-1}$.

Because of the symmetry of the relation "similar", we often say "A and B are similar" instead of saying "A is similar to B".

Similar matrices have a lot in common:

Proposition 2.1.4. Let A and B be two similar matrices. Then:
(a) A and B have the same rank.
(b) A and B have the same nullity.
(c) A and B have the same determinant.
(d) A and B have the same characteristic polynomial.
(e) A and B have the same eigenvalues, with the same algebraic multiplicities and the same geometric multiplicities.
(f) For any $k \in \mathbb{N}$, the matrices A^k and B^k are similar.

Proof. See notes. □

There is a notation $A \sim B$ for "A and B are similar".

2.2. Unitary similarity

Definition 2.2.1. Let A and B be two matrices in $\mathbb{C}^{n\times n}$. We say that A is **unitary similar** to B if there exists a **unitary** matrix $W \in \mathrm{U}_n(\mathbb{C})$ such that $B = WA\underbrace{W^*}_{=W^{-1}}$.

We write "$A \overset{\text{us}}{\sim} B$" for "$A$ is unitary similar to B".

Example 2.2.2. The matrix $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is unitary similar to the matrix $\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, since

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} = W \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} W^* \qquad \text{for } W = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \in \mathrm{U}_2(\mathbb{C}).$$

It is clear that any two unitary similar matrices are similar. The converse is not true (exercise).

"Unitary similar", just like "similar", is an equivalence relation.

2.3. Schur triangularization

We are now ready for one more matrix decomposition, the so-called **Schur triangularization** (aka **Schur decomposition**):

Theorem 2.3.1 (Schur triangularization theorem)**.** Let $A \in \mathbb{C}^{n \times n}$. Then, there exists a unitary matrix $U \in \mathrm{U}_n(\mathbb{C})$ and an upper-triangular matrix $T \in \mathbb{C}^{n \times n}$ such that $A = UTU^*$.

In other words, A is unitary similar to some upper-triangular matrix.

Example 2.3.2. Let $A = \begin{pmatrix} 1 & 3 \\ -3 & 7 \end{pmatrix}$. Then, a Schur triangularization of A is $A = UTU^*$, where

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad \text{and} \qquad T = \begin{pmatrix} 4 & 6 \\ 0 & 4 \end{pmatrix}.$$

Proof. Next time. □