Math 504: Advanced Linear Algebra

Hugo Woerdeman, with edits by Darij Grinberg*

November 15, 2021 (unfinished!)

Contents

Math 504 Lecture 19

1. Hermitian matrices (cont'd)

1.1. Consequences of the interlacing theorem

Recall: If $A \in \mathbb{C}^{n\times n}$ is a Hermitian matrix (i.e., a square matrix satisfying $A^* = A$), then we denote its eigenvalues by $\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A)$ in weakly increasing order (with multiplicities). This makes sense, since we know that these eigenvalues are reals.

Last time, Hugo proved:

Theorem 1.1.1 (Cauchy's interlacing theorem, aka eigenvalue interlacing theorem)**.** Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. Let $j \in [n]$. Let $B \in \mathbb{C}^{(n-1)\times(n-1)}$ be the matrix obtained from A by removing the j-th row and the j-th column. Then,

$$\lambda_1(A) \leq \lambda_1(B) \leq \lambda_2(A) \leq \lambda_2(B) \leq \cdots \leq \lambda_{n-1}(A) \leq \lambda_{n-1}(B) \leq \lambda_n(A).$$

In other words,

$$\lambda_i(A) \leq \lambda_i(B) \leq \lambda_{i+1}(A) \qquad \text{for each } i \in [n-1].$$

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

A converse of this theorem also holds:

Proposition 1.1.2. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ and $\mu_1, \mu_2, \ldots, \mu_{n-1}$ be real numbers satisfying

$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \mu_2 \leq \cdots \leq \lambda_{n-1} \leq \mu_{n-1} \leq \lambda_n.$$

Then, there exist $n-1$ reals $y_1, y_2, \ldots, y_{n-1} \in \mathbb{R}$ and a real $a \in \mathbb{R}$ such that the matrix

$$A := \begin{pmatrix} \mu_1 & & & & y_1 \\ & \mu_2 & & & y_2 \\ & & \ddots & & \vdots \\ & & & \mu_{n-1} & y_{n-1} \\ y_1 & y_2 & \cdots & y_{n-1} & a \end{pmatrix}$$

(where all empty cells are supposed to be filled with 0s) has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. (This matrix is, of course, Hermitian, since it is real symmetric.)

Proof. Omitted. (Exercise?) □

Now, let us derive some consequences from Cauchy's interlacing theorem. We begin with a straightforward generalization:

Corollary 1.1.3 (Cauchy's interlacing theorem for multiple deletions)**.** Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix. Let $r \in \{0, 1, \ldots, n\}$. Let $C \in \mathbb{C}^{r \times r}$ be the result of removing $n-r$ rows and the corresponding $n-r$ columns from A. (That is, we pick some $j_1 < j_2 < \cdots < j_{n-r}$, and we remove the j_1-st, j_2-nd, ..., j_{n-r}-th rows from A, and we remove the j_1-st, j_2-nd, ..., j_{n-r}-th columns from A.) Then, for each $j \in [r]$, we have

$$\lambda_j(A) \leq \lambda_j(C) \leq \lambda_{j+n-r}(A).$$

Proof. Induction on $n-r$.

The *base case* ($n-r=0$) is trivial, since $C = A$ in this case.

In the *induction step*, we obtain C from B by removing a single row and the corresponding column. Thus, by the original Cauchy interlacing theorem, we get $\lambda_j(B) \leq \lambda_j(C)$. However, by the induction hypothesis, we get $\lambda_j(A) \leq \lambda_j(B)$. Combining these inequalities, we get $\lambda_j(A) \leq \lambda_j(C)$. The remaining inequality $\lambda_j(C) \leq \lambda_{j+n-r}(A)$ is proved similarly: By the original Cauchy interlacing theorem, we get $\lambda_j(C) \leq \lambda_{j+1}(B)$. However, by the induction hypothesis, we get $\lambda_{j+1}(B) \leq \lambda_{j+1+(n-r-1)}(A) = \lambda_{j+n-r}(A)$. □

The next corollary provides a minimum/maximum description of the sum of the first m smallest/largest eigenvalues of a Hermitian matrix:

Corollary 1.1.4. Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix. Let $m \in \{0, 1, \ldots, n\}$. Then,

$$\lambda_1(A) + \lambda_2(A) + \cdots + \lambda_m(A) = \min_{\text{isometries } V \in \mathbb{C}^{n \times m}} \operatorname{Tr}(V^* A V)$$

and

$$\lambda_{n-m+1}(A) + \lambda_{n-m+2}(A) + \cdots + \lambda_n(A) = \max_{\text{isometries } V \in \mathbb{C}^{n\times m}} \operatorname{Tr}(V^* A V).$$

Proof. First of all, it suffices to show the first equality, because the second follows by applying the first to $-A$ instead of A.

First, we shall show that

$$\lambda_1(A) + \lambda_2(A) + \cdots + \lambda_m(A) \leq \operatorname{Tr}(V^* A V) \qquad \text{for every isometry } V \in \mathbb{C}^{n\times m}.$$

Indeed, let $V \in \mathbb{C}^{n\times m}$ be an isometry. Thus, V is an $n \times m$-matrix whose columns are orthonormal. As we have seen in the first chapter(?), we can extend each orthonormal tuple of vectors to an orthonormal basis. Doing this to the columns of V, we thus obtain an orthonormal basis of $\mathbb{C}^n$ whose first m entries are the columns of V. Let U be the matrix whose columns are the entries of this basis. Then,

$$U = \begin{pmatrix} V & \widetilde{V} \end{pmatrix} \qquad \text{(in block-matrix notation)}$$

by construction of this basis, and furthermore the matrix U is unitary since its columns form an orthonormal basis.

Since U is unitary, we have $U^* A U \sim A$ and therefore

$$\lambda_j(U^* A U) = \lambda_j(A) \qquad \text{for all } j \in [n].$$

However, $U = \begin{pmatrix} V & \widetilde{V} \end{pmatrix}$ entails

$$U^* A U = \begin{pmatrix} V & \widetilde{V} \end{pmatrix}^* A \begin{pmatrix} V & \widetilde{V} \end{pmatrix} = \begin{pmatrix} V^* \\ \widetilde{V}^* \end{pmatrix} A \begin{pmatrix} V & \widetilde{V} \end{pmatrix} = \begin{pmatrix} V^* A V & * \\ * & * \end{pmatrix},$$

where the three $*$s mean blocks that we don't care about. So the matrix $V^* A V$ is obtained from $U^* A U$ by removing a bunch of rows and the corresponding columns. Hence, the previous corollary yields

$$\lambda_j(U^* A U) \leq \lambda_j(V^* A V) \qquad \text{for all } j \in [m]$$

(since $U^* A U$ is Hermitian (because A is Hermitian)). In other words,

$$\lambda_j(A) \leq \lambda_j(V^* A V) \qquad \text{for all } j \in [m]$$

(since $\lambda_j(U^* A U) = \lambda_j(A)$). Adding these inequalities together, we obtain

$$\begin{aligned}
&\lambda_1(A) + \lambda_2(A) + \cdots + \lambda_m(A) \\
&\leq \lambda_1(V^* A V) + \lambda_2(V^* A V) + \cdots + \lambda_m(V^* A V) \\
&= (\text{the sum of all eigenvalues of } V^* A V) \\
&\qquad \left(\begin{array}{c}\text{since } V^* A V \text{ is an } m \times m\text{-matrix} \\ \text{and thus has } m \text{ eigenvalues}\end{array}\right) \\
&= \operatorname{Tr}(V^* A V)
\end{aligned}$$

(since the sum of all eigenvalues of a matrix is the trace of this matrix).

Now, we need to show that there exists a unitary matrix $V \in \mathbb{C}^{n\times m}$ such that

$$\lambda_1(A) + \lambda_2(A) + \cdots + \lambda_m(A) = \operatorname{Tr}(V^* A V).$$

To do this, we construct V as follows: We pick an eigenvector x_i of A at eigenvalue $\lambda_i(A)$ for each $i \in [n]$ in such a way that $(x_1, x_2, \ldots, x_n)$ is an orthonormal basis of $\mathbb{C}^n$. (This is possible because of Theorem 2.6.1 **(b)**.) Now, let $V \in \mathbb{C}^{n\times m}$ be the matrix whose columns are $x_1, x_2, \ldots, x_m$. This matrix V is an isometry, since $x_1, x_2, \ldots, x_m$ are orthonormal. Moreover,

$$\begin{aligned} V^* A V &= \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_m^* \end{pmatrix} A \begin{pmatrix} x_1 & x_2 & \cdots & x_m \end{pmatrix} \\ &= \begin{pmatrix} x_1^* A x_1 & x_1^* A x_2 & \cdots & x_1^* A x_m \\ x_2^* A x_1 & x_2^* A x_2 & \cdots & x_2^* A x_m \\ \vdots & \vdots & \ddots & \vdots \\ x_m^* A x_1 & x_m^* A x_2 & \cdots & x_m^* A x_m \end{pmatrix}, \end{aligned}$$

so that

$$\begin{aligned} \operatorname{Tr}(V^* A V) &= \sum_{j=1}^{m} x_j^* \underbrace{A x_j}_{\substack{= \lambda_j(A) x_j \\ \text{(since } x_j \text{ is an eigenvector of } A \\ \text{at eigenvalue } \lambda_j(A))}} = \sum_{j=1}^{m} \lambda_j(A) \underbrace{x_j^* x_j}_{\substack{= \left|\left|x_j\right|\right|^2 = 1 \\ \text{(since } (x_1, x_2, \ldots, x_n) \\ \text{is an orthonormal basis)}}} \\ &= \sum_{j=1}^{m} \lambda_j(A) = \lambda_1(A) + \lambda_2(A) + \cdots + \lambda_m(A). \end{aligned}$$

This is precisely what we needed. Thus, we conclude that

$$\lambda_1(A) + \lambda_2(A) + \cdots + \lambda_m(A) = \min_{\text{isometries } V \in \mathbb{C}^{n\times m}} \operatorname{Tr}(V^* A V).$$

As we said above, this completes the proof. □

Corollary 1.1.5. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian $n \times n$-matrix. Let $m \in \{0, 1, \ldots, n\}$. Let $i_1, i_2, \ldots, i_m \in [n]$ be distinct. Then,

$$\begin{aligned} \lambda_1(A) + \lambda_2(A) + \cdots + \lambda_m(A) &\leq A_{i_1,i_1} + A_{i_2,i_2} + \cdots + A_{i_m,i_m} \\ &\leq \lambda_{n-m+1}(A) + \lambda_{n-m+2}(A) + \cdots + \lambda_n(A). \end{aligned}$$

In words: For a Hermitian matrix A, each sum of m distinct diagonal entries of A is sandwiched between the sum of the m smallest eigenvalues of A and the sum of the m largest eigenvalues of A.

Proof. Let C be the matrix obtained from A by removing all but the i_1-st, i_2-nd, ..., i_m-th rows and the corresponding columns of A. Then,

$$\operatorname{Tr} C = A_{i_1,i_1} + A_{i_2,i_2} + \cdots + A_{i_m,i_m}.$$

However, Cauchy's interlacing theorem for multiple deletions yields

$$\lambda_j(A) \leq \lambda_j(C) \qquad \text{for each } j \in [m].$$

Summing these up, we obtain

$$\begin{aligned}\lambda_1(A) + \lambda_2(A) + \cdots + \lambda_m(A) &\leq \lambda_1(C) + \lambda_2(C) + \cdots + \lambda_m(C) \\ &= (\text{the sum of all eigenvalues of } C) \\ &= \operatorname{Tr} C = A_{i_1,i_1} + A_{i_2,i_2} + \cdots + A_{i_m,i_m}.\end{aligned}$$

So we have proved the first of the required two inequalities. The second follows by applying the first to $-A$ instead of A. □

The above corollary has a bunch of consequences that are obtained by restating it in terms of something called **majorization**. Let us define this concept and see what it entails.

1.2. Introduction to majorization theory

Convention 1.2.1. Let $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ be a column vector with real entries. Then, for each $i \in [n]$, we let $x_i^\downarrow$ denote the i-th largest entry of x. So $\left(x_1^\downarrow, x_2^\downarrow, \ldots, x_n^\downarrow\right)$ is the unique permutation of the tuple $(x_1, x_2, \ldots, x_n)$ that satisfies

$$x_1^\downarrow \geq x_2^\downarrow \geq \cdots \geq x_n^\downarrow.$$

For example, if $x = (3,5,2)^T$, then $x_1^\downarrow = 5$ and $x_2^\downarrow = 3$ and $x_3^\downarrow = 2$.

Similarly, we define $x_i^\uparrow$ to be the i-th smallest entry of x.

Definition 1.2.2. Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ be two column vectors with real entries. Then, we say that x **majorizes** y (and we write $x \succcurlyeq y$) if and only if we have

$$\sum_{i=1}^{m} x_i^\downarrow \geq \sum_{i=1}^{m} y_i^\downarrow \qquad \text{for each } m \in [n],$$

with equality for $m = n$ (and possibly for other m's). In other words, x majorizes

y if and only if

$$\begin{aligned} x_1^\downarrow &\geq y_1^\downarrow; \\ x_1^\downarrow + x_2^\downarrow &\geq y_1^\downarrow + y_2^\downarrow; \\ x_1^\downarrow + x_2^\downarrow + x_3^\downarrow &\geq y_1^\downarrow + y_2^\downarrow + y_3^\downarrow; \\ &\ldots; \\ x_1^\downarrow + x_2^\downarrow + \cdots + x_{n-1}^\downarrow &\geq y_1^\downarrow + y_2^\downarrow + \cdots + y_{n-1}^\downarrow; \\ x_1^\downarrow + x_2^\downarrow + \cdots + x_n^\downarrow &= y_1^\downarrow + y_2^\downarrow + \cdots + y_n^\downarrow. \end{aligned}$$

Example 1.2.3. We have

$$\begin{pmatrix} 1 \\ 3 \\ 5 \\ 7 \end{pmatrix} \geq \begin{pmatrix} 2 \\ 2 \\ 6 \\ 6 \end{pmatrix},$$

since

$$\begin{aligned} 7 &\geq 6; \\ 7+5 &\geq 6+6; \\ 7+5+3 &\geq 6+6+2; \\ 7+5+3+1 &= 6+6+2+2. \end{aligned}$$

Example 1.2.4. We don't have

$$\begin{pmatrix} 1 \\ 3 \\ 5 \\ 7 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 2 \\ 6 \\ 8 \end{pmatrix},$$

since we don't have $7 \geq 8$.

The intuition behind majorization is the following: x majorizes y if and only if you can obtain y from x by "having the entries come closer together (while keeping the average equal)".

Proposition 1.2.5. Majorization is a partial order: i.e., a reflexive, antisymmetric and transitive relation.

However, it is not a total order: For example, the sequences

$$x := \begin{pmatrix} 2 \\ 2 \\ 4 \\ 6 \end{pmatrix} \qquad \text{and} \qquad y = \begin{pmatrix} 1 \\ 3 \\ 5 \\ 5 \end{pmatrix}$$

satisfy neither $x \succcurlyeq y$ nor $y \succcurlyeq x$, since we have $6 > 5$ but $6 + 4 + 2 < 5 + 5 + 3$.

Now we can restate our last corollary as follows:

Corollary 1.2.6 (Schur's theorem)**.** Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian $n \times n$-matrix. Then,

$$\left(A_{1,1}, A_{2,2}, \ldots, A_{n,n}\right)^T \succcurlyeq \left(\lambda_1(A), \lambda_2(A), \ldots, \lambda_n(A)\right)^T.$$

In words: The tuple of diagonal entries of A majorizes the tuple of eigenvalues of A.

Proof. We need to show that

- the sum of the m largest diagonal entries of A is $\geq$ to the sum of the m largest eigenvalues of A for each $m \in [n]$;
- the sum of all diagonal entries of A equals the sum of all eigenvalues of A.

But the first of these two statements is the first inequality in the above corollary, whereas the second statement is the well-known theorem that the trace of a matrix is the sum of its eigenvalues. □

Now, what can we do with majorizing vectors? Here is probably the most important property:

Theorem 1.2.7 (Karamata's inequality)**.** Let $I \subseteq \mathbb{R}$ be an interval. Let $f : I \to \mathbb{R}$ be a convex function. Let $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ and $y = (y_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n$ be two vectors such that $x \succcurlyeq y$. Then,

$$f(x_1) + f(x_2) + \cdots + f(x_n) \geq f(y_1) + f(y_2) + \cdots + f(y_n).$$

For example, applying this to $f(t) = t^2$, we obtain

$$x_1^2 + x_2^2 + \cdots + x_n^2 \geq y_1^2 + y_2^2 + \cdots + y_n^2.$$

Next time, we will prove Karamata's inequality.