Math 504: Advanced Linear Algebra

Hugo Woerdeman, with edits by Darij Grinberg*

November 19, 2021 (unfinished!)

Contents

1. Hermitian matrices (cont'd) **1**
1.1. Introduction to majorization theory (cont'd) 1

Math 504 Lecture 21

1. Hermitian matrices (cont'd)

1.1. Introduction to majorization theory (cont'd)

Recall:

- If $x \in \mathbb{R}^n$ is a column vector and $i \in [n]$, then x_i denotes the i-th coordinate (= entry) of x.
- If $x \in \mathbb{R}^n$ is a column vector, then $x^{\downarrow}$ denotes the weakly decreasing permutation of x (that is, the vector with the same entries as x but in weakly decreasing order).
- For two vectors $x, y \in \mathbb{R}^n$, we say that x **majorizes** y (and we write $x \succcurlyeq y$) if and only if

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

  and

  $$x_1 + x_2 + \cdots + x_n = y_1 + y_2 + \cdots + y_n.$$

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

- A **Robin Hood move** (short **RH move**) transforms a vector $x \in \mathbb{R}^n$ into a vector $y \in \mathbb{R}^n$ that is constructed as follows: Pick two distinct $i, j \in [n]$ satisfying $x_i \leq x_j$, and pick two numbers $u, v \in [x_i, x_j]$ satisfy $u + v = x_i + x_j$, and replace the i-th and the j-th entries of x by u and v.

  This RH move is said to be an **order-preserving RH move** (short **OPRH move**) if both x and y are weakly decreasing.

Last time we proved:

Theorem 1.1.1 (RH criterion for majorization)**.** Let $x, y \in \mathbb{R}^n$ be two weakly decreasing column vectors. Then, $x \succcurlyeq y$ if and only if y can be obtained from x by a (finite) sequence of OPRH moves.

Now, what can we do with majorization? Probably the most important property of majorizing pairs of vectors is the following:

Theorem 1.1.2 (Karamata's inequality)**.** Let $I \subseteq \mathbb{R}$ be an interval. Let $f : I \to \mathbb{R}$ be a convex function. Let $x \in I^n$ and $y \in I^n$ be two vectors such that $x \succcurlyeq y$. Then,

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

Before we prove this, let us state a simple corollary:

Corollary 1.1.3 (Jensen's inequality)**.** Let $I \subseteq \mathbb{R}$ be an interval. Let $f : I \to \mathbb{R}$ be a convex function. Let $x_1, x_2, \ldots, x_n \in I$. Let $m = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$. Then,

$$f(x_1) + f(x_2) + \cdots + f(x_n) \geq nf(m).$$

Proof. This follows from Karamata's inequality, since it is easy to see (exercise) that $(x_1, x_2, \ldots, x_n)^T \succcurlyeq (m, m, \ldots, m)^T$. □

Let us now prove Karamata's inequality:

Proof of Karamata's inequality. It is enough to prove the claim in the case when x and y are weakly decreasing (because permuting the entries of any of x and y does not change anything).

Furthermore, it is enough to prove the claim in the case when $x \overset{\text{OPRH}}{\longrightarrow} y$ (this means that y is obtained from x by a single OPRH move). In fact, if we can show this, then it will mean that the sum $f(x_1) + f(x_2) + \cdots + f(x_n)$ decreases (weakly) every time we apply an OPRH move. Therefore, if y is obtained from x by a sequence of OPRH moves, then $f(x_1) + f(x_2) + \cdots + f(x_n) \geq f(y_1) + f(y_2) + \cdots + f(y_n)$. But the theorem we proved last time shows that this is always satisfied when $x \succcurlyeq y$.

So let us assume that $x \overset{\text{OPRH}}{\longrightarrow} y$. Thus, y is obtained from x by picking two entries x_i and x_j with $x_i \leq x_j$ and replacing them by u and v, where $u, v \in \left[x_i, x_j\right]$ with $u + v = x_i + x_j$. Consider these x_i, x_j, u, v. It clearly suffices to show that

$$f(x_i) + f\left(x_j\right) \geq f(u) + f(v).$$

How do we do this? From $u \in \left[x_i, x_j\right]$, we obtain

$$u = \lambda x_i + (1 - \lambda) x_j \qquad \text{for some } \lambda \in [0, 1]$$

(namely, $\lambda = \dfrac{1}{x_i - x_j}\left(u - x_j\right)$). Consider this λ. Then,

$$v = (1 - \lambda) x_i + \lambda x_j \qquad \left(\text{since } u + v = x_i + x_j\right).$$

However, f is convex. From $u = \lambda x_i + (1 - \lambda) x_j$, we obtain

$$f(u) = f\left(\lambda x_i + (1 - \lambda) x_j\right) \leq \lambda f(x_i) + (1 - \lambda) f\left(x_j\right) \qquad (\text{since } f \text{ is convex})$$

and

$$f(v) = f\left((1 - \lambda) x_i + \lambda x_j\right) \leq (1 - \lambda) f(x_i) + \lambda f\left(x_j\right).$$

Adding together these two inequalities, we obtain

$$\begin{aligned} f(u) + f(v) &\leq \left(\lambda f(x_i) + (1 - \lambda) f\left(x_j\right)\right) + \left((1 - \lambda) f(x_i) + \lambda f\left(x_j\right)\right) \\ &= f(x_i) + f\left(x_j\right), \qquad \text{qed.} \end{aligned}$$

So Karamata's inequality is proved. □

Karamata's inequality has lots of applications, since there are many convex functions around. For instance:

- $f(t) = t^n$ defines a convex function on $\mathbb{R}$ whenever $n \in \mathbb{N}$ is even.
- $f(t) = t^n$ defines a convex function on $\mathbb{R}_+$ whenever $n \in \mathbb{R} \setminus (0, 1)$. Otherwise, it defines a concave function (so that $-f$ is a convex function).
- $f(t) = \sin t$ defines a concave function on $[0, \pi]$ and a convex function on $[\pi, 2\pi]$.

Karamata's inequality has a converse: If $x, y \in \mathbb{R}^n$ are two vectors such that

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

for every convex function $f : \mathbb{R} \to \mathbb{R}$, then $x \succcurlyeq y$. Even better:

Theorem 1.1.4 (absolute-value criterion for majorization)**.** Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ be two vectors. Then, $x \succcurlyeq y$ if and only if all $t \in \mathbb{R}$ satisfy

$$|x_1 - t| + |x_2 - t| + \cdots + |x_n - t| \geq |y_1 - t| + |y_2 - t| + \cdots + |y_n - t|.$$

Proof. $\Longrightarrow$: Assume that $x \succcurlyeq y$. Let $t \in \mathbb{R}$. Consider the function

$$f_t : \mathbb{R} \to \mathbb{R},$$
$$z \mapsto |z - t|.$$

This function f_t is convex (this follows easily from the triangle inequality). Hence, Karamata's inequality yields

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

By the definition of f_t, this means

$$|x_1 - t| + |x_2 - t| + \cdots + |x_n - t| \geq |y_1 - t| + |y_2 - t| + \cdots + |y_n - t|.$$

So we have proved the "$\Longrightarrow$" direction of the theorem.

$\Longleftarrow$: We shall refer to the inequality

$$|x_1 - t| + |x_2 - t| + \cdots + |x_n - t| \geq |y_1 - t| + |y_2 - t| + \cdots + |y_n - t|$$

as the **absolute value inequality**. So we assume that the absolute value inequality holds for all $t \in \mathbb{R}$. (Actually, it will suffice to assume that it holds for all $t \in \{x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n\}$.)

We must prove that $x \succcurlyeq y$.

WLOG assume that x and y are weakly decreasing (since permuting the entries changes neither the absolute value inequality nor the claim $x \succcurlyeq y$).

Let $t = \max\{x_1, y_1\}$. Then, for all $i \in [n]$, we have $t = \max\{x_1, y_1\} \geq x_1 \geq x_i$ (since x is weakly decreasing) and therefore $|x_i - t| = t - x_i$ and similarly $|y_i - t| = t - y_i$. Thus, the absolute value inequality rewrites as

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

In other words,

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

In other words,

$$x_1 + x_2 + \cdots + x_n \leq y_1 + y_2 + \cdots + y_n.$$

Similarly, by taking $t = \min\{x_n, y_n\}$, we obtain

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

Combining these two inequalities, we obtain

$$x_1 + x_2 + \cdots + x_n = y_1 + y_2 + \cdots + y_n.$$

Now, let $k \in [n]$. Set $t = x_k$. Then, since x is weakly decreasing, we have

$$x_1 \geq x_2 \geq \cdots \geq x_k = t \geq x_{k+1} \geq x_{k+2} \geq \cdots \geq x_n.$$

Thus,

$$\begin{aligned}
&|x_1 - t| + |x_2 - t| + \cdots + |x_n - t| \\
&= \left(\underbrace{|x_1 - t|}_{= x_1 - t} + \underbrace{|x_2 - t|}_{= x_2 - t} + \cdots + \underbrace{|x_k - t|}_{= x_k - t} \right) + \left(\underbrace{|x_{k+1} - t|}_{= t - x_{k+1}} + \underbrace{|x_{k+2} - t|}_{= t - x_{k+2}} + \cdots + \underbrace{|x_n - t|}_{= t - x_n} \right) \\
&= ((x_1 - t) + (x_2 - t) + \cdots + (x_k - t)) + ((t - x_{k+1}) + (t - x_{k+2}) + \cdots + (t - x_n)) \\
&= (x_1 + x_2 + \cdots + x_k) - kt + (n - k)\, t - (x_{k+1} + x_{k+2} + \cdots + x_n) \\
&= 2\,(x_1 + x_2 + \cdots + x_k) + (n - 2k)\, t - (x_1 + x_2 + \cdots + x_n)
\end{aligned}$$

(here, I have added $x_1 + x_2 + \cdots + x_k$ to the first parenthesis and subtracted it back from the last) and

$$\begin{aligned}
&|y_1 - t| + |y_2 - t| + \cdots + |y_n - t| \\
&= \left(\underbrace{|y_1 - t|}_{\geq y_1 - t} + \underbrace{|y_2 - t|}_{\geq y_2 - t} + \cdots + \underbrace{|y_k - t|}_{\geq y_k - t} \right) + \left(\underbrace{|y_{k+1} - t|}_{\geq t - y_{k+1}} + \underbrace{|y_{k+2} - t|}_{\geq t - y_{k+2}} + \cdots + \underbrace{|y_n - t|}_{\geq t - y_n} \right) \\
&\geq ((y_1 - t) + (y_2 - t) + \cdots + (y_k - t)) + ((t - y_{k+1}) + (t - y_{k+2}) + \cdots + (t - y_n)) \\
&= (y_1 + y_2 + \cdots + y_k) - kt + (n - k)\, t - (y_{k+1} + y_{k+2} + \cdots + y_n) \\
&= 2\,(y_1 + y_2 + \cdots + y_k) + (n - 2k)\, t - (y_1 + y_2 + \cdots + y_n).
\end{aligned}$$

Now, the absolute value inequality

$$|x_1 - t| + |x_2 - t| + \cdots + |x_n - t| \geq |y_1 - t| + |y_2 - t| + \cdots + |y_n - t|$$

becomes

$$\begin{aligned}
&2\,(x_1 + x_2 + \cdots + x_k) + (n - 2k)\, t - (x_1 + x_2 + \cdots + x_n) \\
&\geq |y_1 - t| + |y_2 - t| + \cdots + |y_n - t| \\
&\geq 2\,(y_1 + y_2 + \cdots + y_k) + (n - 2k)\, t - (y_1 + y_2 + \cdots + y_n).
\end{aligned}$$

Subtracting $(n - 2k)\, t$ from both sides, we obtain

$$\begin{aligned}
&2\,(x_1 + x_2 + \cdots + x_k) - (x_1 + x_2 + \cdots + x_n) \\
&\geq 2\,(y_1 + y_2 + \cdots + y_k) - (y_1 + y_2 + \cdots + y_n).
\end{aligned}$$

Adding the equality

$$x_1 + x_2 + \cdots + x_n = y_1 + y_2 + \cdots + y_n$$

to this inequality, we obtain

$$2(x_1 + x_2 + \cdots + x_k) \geq 2(y_1 + y_2 + \cdots + y_k).$$

In other words,

$$x_1 + x_2 + \cdots + x_k \geq y_1 + y_2 + \cdots + y_k.$$

Since x and y are weakly decreasing, this shows that $x \succcurlyeq y$ (since $x_1 + x_2 + \cdots + x_n = y_1 + y_2 + \cdots + y_n$). This proves the "$\Longleftarrow$" direction of the theorem. □

Majorizing pairs of vectors are closely related to **doubly stochastic matrices**:

Definition 1.1.5. A matrix $S \in \mathbb{R}^{n \times n}$ is said to be **doubly stochastic** if its entries $S_{i,j}$ satisfy the following three conditions:

1. We have $S_{i,j} \geq 0$ for all i, j.
2. We have $\sum\limits_{j=1}^{n} S_{i,j} = 1$ for each $i \in [n]$.
3. We have $\sum\limits_{i=1}^{n} S_{i,j} = 1$ for each $j \in [n]$.

In other words, a doubly stochastic matrix is an $n \times n$-matrix whose entries are nonnegative reals and whose rows and columns have sum 1 each.

Exercise 1.1.1. Show that even if we allow S to be rectangular, the conditions 2 and 3 still force S to be a square matrix.

Example 1.1.6. **(a)** The matrix $\begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{5}{12} & \frac{7}{12} \end{pmatrix}$ is doubly stochastic.

(b) Each doubly stochastic 2×2-matrix has the form

$$\begin{pmatrix} \lambda & 1-\lambda \\ 1-\lambda & \lambda \end{pmatrix} \qquad \text{for some } \lambda \in [0,1].$$

(c) Any permutation matrix is doubly stochastic.

Proposition 1.1.7. Let $S \in \mathbb{R}^{n \times n}$ be a matrix whose entries are nonnegative reals. Let $e = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$. Then, S is doubly stochastic if and only if $Se = e$ and $e^T S = e^T$.

Corollary 1.1.8. Any product of doubly stochastic matrices is again doubly stochastic.

Exercise 1.1.2. Prove this proposition and this corolllary.

Now, the connection with majorization:

Theorem 1.1.9. Let $x, y \in \mathbb{R}^n$ be two vectors. Then, $x \succcurlyeq y$ if and only if there exists a doubly stochastic matrix S such that $y = Sx$.

Proof. $\Longrightarrow$: Assume that $x \succcurlyeq y$. We must prove that there exists a doubly stochastic matrix S such that $y = Sx$.

By example **(c)** (and the corollary), it suffices to show this in the case when x and y are weakly decreasing.

By the corollary, it suffices to show this in the case when $x \overset{\text{OPRH}}{\longrightarrow} y$ (since in the general case, y is obtained from x by a sequence of such moves).

So let us assume that $x \overset{\text{OPRH}}{\longrightarrow} y$. Thus, y is obtained from x by picking two entries x_i and x_j with $x_i \leq x_j$ and replacing them by u and v, where $u, v \in \left[x_i, x_j\right]$ with $u + v = x_i + x_j$. Consider these x_i, x_j, u, v. It clearly suffices to show that

$$f(x_i) + f\left(x_j\right) \geq f(u) + f(v).$$

How do we do this? From $u \in \left[x_i, x_j\right]$, we obtain

$$u = \lambda x_i + (1 - \lambda) x_j \qquad \text{for some } \lambda \in [0, 1]$$

(namely, $\lambda = \dfrac{1}{x_i - x_j}\left(u - x_j\right)$). Consider this λ. Then,

$$v = (1 - \lambda) x_i + \lambda x_j \qquad \left(\text{since } u + v = x_i + x_j\right).$$

This entails that $y = Sx$, where S is a matrix given by

$$\begin{aligned} &S_{i,i} = \lambda, \qquad S_{i,j} = 1 - \lambda, \qquad S_{j,i} = 1 - \lambda, \qquad S_{j,j} = \lambda, \\ &S_{k,k} = 1 \qquad \text{for each } k \notin \{i, j\}, \\ &S_{k,\ell} = 0 \qquad \text{for all remaining } k, \ell. \end{aligned}$$

For example, if $i = 2$ and $j = 4$, then

$$S = \begin{pmatrix} 1 & & & \\ & \lambda & & 1 - \lambda \\ & & 1 & \\ & 1 - \lambda & & \lambda \end{pmatrix}$$

(where all empty cells are filled with zeroes). In this case,

$$Sx = S \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 \\ \lambda x_2 + (1-\lambda) x_4 \\ x_3 \\ (1-\lambda) x_2 + \lambda x_4 \end{pmatrix} = \begin{pmatrix} x_1 \\ u \\ x_3 \\ v \end{pmatrix} = y.$$

So we are done proving the "$\Longrightarrow$" direction.

$\Longleftarrow$: Assume that $y = Sx$ for some doubly stochastic S. Then,

$$y_i = (Sx)_i = \sum_{j=1}^{n} S_{i,j} x_j \qquad \text{for every } i \in [n].$$

Thus, for every convex function $f : \mathbb{R} \to \mathbb{R}$, we have

$$\begin{aligned}
& f(y_1) + f(y_2) + \cdots + f(y_n) \\
&= \underbrace{f\left(\sum_{j=1}^{n} S_{1,j} x_j\right)}_{\substack{\leq \sum_{j=1}^{n} S_{1,j} f(x_j) \\ \text{(by the weighted Jensen inequality,} \\ \text{since the } S_{1,j}\text{'s are nonnegative reals} \\ \text{with sum } =1)}} + \underbrace{f\left(\sum_{j=1}^{n} S_{2,j} x_j\right)}_{\substack{\leq \sum_{j=1}^{n} S_{2,j} f(x_j) \\ \text{(by the weighted Jensen inequality,} \\ \text{since the } S_{2,j}\text{'s are nonnegative reals} \\ \text{with sum } =1)}} + \cdots + \underbrace{f\left(\sum_{j=1}^{n} S_{n,j} x_j\right)}_{\substack{\leq \sum_{j=1}^{n} S_{n,j} f(x_j) \\ \text{(by the weighted Jensen inequality,} \\ \text{since the } S_{n,j}\text{'s are nonnegative reals} \\ \text{with sum } =1)}} \\
&\leq \sum_{j=1}^{n} S_{1,j} f(x_j) + \sum_{j=1}^{n} S_{2,j} f(x_j) + \cdots + \sum_{j=1}^{n} S_{n,j} f(x_j) \\
&= \sum_{j=1}^{n} \underbrace{(S_{1,j} + S_{2,j} + \cdots + S_{n,j})}_{\substack{=1 \\ \text{(since } S \text{ is doubly stochastic)}}} f(x_j) \\
&= \sum_{j=1}^{n} f(x_j) = f(x_1) + f(x_2) + \cdots + f(x_n).
\end{aligned}$$

In particular, we can apply this to the convex function $f_t : \mathbb{R} \to \mathbb{R},\ z \mapsto |z - t|$ for any $t \in \mathbb{R}$, and we obtain

$$|y_1 - t| + |y_2 - t| + \cdots + |y_n - t| \leq |x_1 - t| + |x_2 - t| + \cdots + |x_n - t|.$$

In other words,

$$|x_1 - t| + |x_2 - t| + \cdots + |x_n - t| \geq |y_1 - t| + |y_2 - t| + \cdots + |y_n - t|.$$

Therefore, by the absolute value criterion, we have $x \succcurlyeq y$. This proves the "$\Longleftarrow$" direction.

□