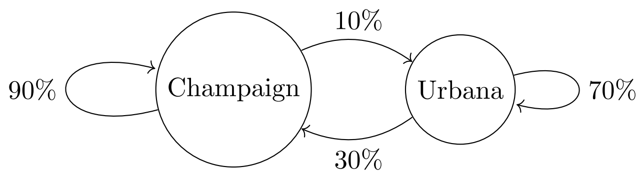
Math 416 Lecture Note: Week 11

Daesung Kim

1 Introduction to Markov Chains

As an application of the diagonalization, we discuss the limiting distribution of Markov chains.

Example 1.1. Suppose that the populations of Champaign and Urbana in 2019 are 90k and 40k. Suppose that every year, it is expected that 90% of people in Champaign stay in the city, the rest move to Urbana, and 70% of people in Urbana stay and the rest move to Champaign. (Assume that there is no movement from outside.)



The question is what the populations of the cities will be after 5 (or 10, 15, and n) years. Let X_n be a column vector that represents the populations of the cities after n years (in 10k), that is,

$$X_n = \begin{pmatrix} \text{the population of Champaign after } n \text{ years} \\ \text{the population of Urbana after } n \text{ years} \end{pmatrix}.$$

For example,

$$X_0 = \begin{pmatrix} 9 \\ 4 \end{pmatrix}.$$

Our goal is to find X_n for all n. Let's find X_1 first. It follows from the assumption that

$$\begin{aligned} (\text{Champaign when } n = 1) &= (\text{Champaign when } n = 0) \times 90\% + (\text{Urbana when } n = 0) \times 30\% \\ (\text{Urbana when } n = 1) &= (\text{Champaign when } n = 0) \times 10\% + (\text{Urbana when } n = 0) \times 70\%. \end{aligned}$$

We can simply write it as

$$\begin{pmatrix} \text{Champaign when } n = 1 \\ \text{Urbana when } n = 1 \end{pmatrix} = \begin{pmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{pmatrix} \begin{pmatrix} \text{Champaign when } n = 0 \\ \text{Urbana when } n = 0 \end{pmatrix},$$

that is, $X_1 = AX_0$ where

$$A = \begin{pmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{pmatrix}.$$

It is easy to see that $X_2 = AX_1$, $X_3 = AX_2$, and so on. In general, we have

$$X_n = AX_{n-1} = A^2 X_{n-2} = \cdots = A^n X_0.$$

Thus, we boil it down to understand A^n for all integer $n \geq 1$. To this end, we diagonalize the matrix A. The characteristic polynomial is

$$f(t) = \det(A - tI) = \det \begin{pmatrix} 0.9 - t & 0.3 \\ 0.1 & 0.7 - t \end{pmatrix} = (t - 0.9)(t - 0.7) - 0.03 = t^2 - 1.6t + 0.6 = (t - 1)(t - 0.6)$$

and so the eigenvalues are $\lambda = 1, 0, 6$. To find the eigenspaces, we observe that

$$A - I = \begin{pmatrix} -0.1 & 0.3 \\ 0.1 & -0.3 \end{pmatrix} \to \begin{pmatrix} 1 & -3 \\ 0 & 0 \end{pmatrix}, \qquad A - 0.6I = \begin{pmatrix} 0.3 & 0.3 \\ 0.1 & 0.1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},$$

which yields that $E_1 = \{t(3, 1) : t \in \mathbb{R}\}$ and $E_{0.6} = \{t(1, -1) : t \in \mathbb{R}\}$. Thus,

$$\begin{aligned} A &= QDQ^{-1} \\ &= \begin{pmatrix} 1 & 3 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0.6 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ -1 & 1 \end{pmatrix}^{-1} \\ &= \frac{1}{4} \begin{pmatrix} 1 & 3 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0.6 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -3 \\ 1 & 1 \end{pmatrix}. \end{aligned}$$

By the previous HW, we have

$$\begin{aligned} A^n &= QD^nQ^{-1} \\ &= \frac{1}{4} \begin{pmatrix} 1 & 3 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} (0.6)^n & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -3 \\ 1 & 1 \end{pmatrix} \\ &= \frac{1}{4} \begin{pmatrix} (0.6)^n + 3 & -3(0.6)^n + 3 \\ -(0.6)^n + 1 & 3(0.6)^n + 1 \end{pmatrix}. \end{aligned}$$

Therefore, we finally find the population after t years

$$X_n = A^n X_0 = \frac{1}{4} \begin{pmatrix} (0.6)^n + 3 & -3(0.6)^n + 3 \\ -(0.6)^n + 1 & 3(0.6)^n + 1 \end{pmatrix} \begin{pmatrix} 9 \\ 4 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 39 - 3(0.6)^n \\ 13 + 3(0.6)^n \end{pmatrix}.$$

Remark 1.2. We note the following.

(i) The sum of each column of A is 1.

(ii) The eigenvalues of A are less than or equal to 1.

(iii) One of the eigenvalues of A is 1.

(iv) The sum of the entries of X_n does not depend on n, only on X_0.

(v) As n tends to ∞, X_n is getting close to $\begin{pmatrix} 39/4 \\ 13/4 \end{pmatrix}$ and

$$\lim_{n\to\infty} A^n = \lim_{n\to\infty} \frac{1}{4} \begin{pmatrix} (0.6)^n + 3 & -3(0.6)^n + 3 \\ -(0.6)^n + 1 & 3(0.6)^n + 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 3 & 3 \\ 1 & 1 \end{pmatrix} := L.$$

(vi) The vector $(3, 1)$ is an eigenvector of A corresponding to 1.

(vii) The limiting matrix L has the property

$$AL = \frac{1}{4} \begin{pmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{pmatrix} \begin{pmatrix} 3 & 3 \\ 1 & 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 3 & 3 \\ 1 & 1 \end{pmatrix} = L.$$

Roughly speaking, a Markov chain is a random model that "the future state depends not on all the previous history but only on the current state." In the example, the expected population in 2020 is determined by the population in 2019. This prediction does not rely on the population before 2018. In other words, it is enough to know the population in 2019, to predict that in 2020.

Example 1.3 (Coupon collector)**.** Suppose that a company issues 3 different types of coupons and there is a collector who wants to collect them all. Assume that each coupon is equally likely to be each of the 3 types. A question is how many times he needs to obtain coupons to get a complete set.

Let C_n be the number of the types he gets after obtaining n coupons, then $C_n \in \{0,1,2,3\}$. We write the probabilities of C_n as a column vector,

$$X_n = \begin{pmatrix} \text{Probability of } C_n = 0 \\ \text{Probability of } C_n = 1 \\ \text{Probability of } C_n = 2 \\ \text{Probability of } C_n = 3 \end{pmatrix}.$$

Note that

$$X_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Suppose $p_i \geq 0$ for $i = 0,1,2,3$, $\sum_{i=0}^{3} p_i = 1$, and

$$X_n = \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix}.$$

Then, $X_{n+1} = (q_0, q_1, q_2, q_3)^t$ and X_n have a relation

$$\begin{aligned} q_0 &= 0 \\ q_1 &= p_0 + \frac{1}{3}p_1 \\ q_2 &= \frac{2}{3}p_1 + \frac{1}{3}p_2 \\ q_3 &= \frac{1}{3}p_2 + \frac{2}{3}p_3 \end{aligned}$$

or

$$X_{n+1} = \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & \frac{2}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 1 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} = AX_n$$

Again, it is enough to understand A^n. By diagonalizing A, we get

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & \frac{2}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -3 & -2 & -1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -3 & -2 & -1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} = QDQ^{-1}.$$

Thus,

$$\begin{aligned} A^n = QD^nQ^{-1} &= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -3 & -2 & -1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \left(\frac{1}{3}\right)^n & 0 & 0 \\ 0 & 0 & \left(\frac{2}{3}\right)^n & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -3 & -2 & -1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 3\left(\frac{1}{3}\right)^n & \left(\frac{1}{3}\right)^n & 0 & 0 \\ 3\left(\frac{2}{3}\right)^n - 6\left(\frac{1}{3}\right)^n & 2\left(\frac{2}{3}\right)^n - 2\left(\frac{1}{3}\right)^n & \left(\frac{2}{3}\right)^n & 0 \\ 1 + 3\left(\frac{1}{3}\right)^n - 3\left(\frac{2}{3}\right)^n & 1 + \left(\frac{1}{3}\right)^n - 2\left(\frac{2}{3}\right)^n & 1 - \left(\frac{2}{3}\right)^n & 1 \end{pmatrix} \end{aligned}$$

and

$$X_n = A^n X_0 = \begin{pmatrix} 0 \\ 3\left(\frac{1}{3}\right)^n \\ 3\left(\frac{2}{3}\right)^n - 6\left(\frac{1}{3}\right)^n \\ 1 + 3\left(\frac{1}{3}\right)^n - 3\left(\frac{2}{3}\right)^n . \end{pmatrix}$$

2 Convergence of Markov Chains

Definition 2.1. A probability vector is a column vector $(a_1, \cdots, a_n)^t \in \mathbb{R}^n$ such that $a_i \geq 0$ and $\sum_{i=1}^n a_i = 1$. A transition matrix is a matrix $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ such that $A_{ij} \geq 0$ and $\sum_{i=1}^n A_{ij} = 1$ for each $j = 1, 2, \cdots, n$.

Theorem 2.2. *Suppose that $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is a transition matrix and there exists $d \geq 1$ such that all the entries of A^d are positive. Then,*

(i) *1 is an eigenvalue for A and the corresponding eigenspace E_1 is*

$$E_1 = \mathrm{Span}(\{u\})$$

where u is a probability vector.

(ii) *If λ is an eigenvalue for A and $\lambda \neq 1$, then $|\lambda| < 1$.*

(iii) *If $u = (u_1, \cdots, u_n)^t$ is the probability vector in (i), then*

$$\lim_{n\to\infty} A^n = \begin{pmatrix} | & | & \cdots & | \\ u & u & \cdots & u \\ | & | & \cdots & | \end{pmatrix} = \begin{pmatrix} u_1 & u_1 & \cdots & u_1 \\ \vdots & \vdots & \vdots & \vdots \\ u_n & u_n & \cdots & u_n \end{pmatrix}.$$

Remark 2.3. How can we interpret the assumption that "all the entries of A^d are positive"? If it is so, the corresponding Markov chain can visit from one state to another state within d steps with positive probability. In other words, it means that all the states are well-mixed by the Markov chain. For example, if $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, then $1, -1$ are the eigenvalues of A, (which violates (ii)) and

$$A^n = \begin{cases} I, & \text{if } n \text{ is even,} \\ A, & \text{if } n \text{ is odd} \end{cases}$$

(which violates (iii)). From now on, we simply assume that $d = 1$.

Theorem 2.4. *If A is a transition matrix, then 1 is an eigenvalue of A and $|\lambda| \leq 1$ for any eigenvalues λ of A.*

Proof. Note that λ is an eigenvalue of A if and only if λ is an eigenvalue of A^t (HW). Let $u = (1, 1, \cdots, 1)^t$, then it is easy to see that $A^t u = u$, which implies that 1 is an eigenvalue of A. (Note that u is not an eigenvector of A.) Let λ be an eigenvalue of A (and A^t) and $v = (v_1, \cdots, v_k)^t$ the corresponding eigenvector of A^t (not A). Suppose that $v_k \geq v_i$ for all $i = 1, 2, \cdots, n$ for some k. Then,

$$(A^t v)_k = (\lambda v)_k = \lambda v_k = \sum_{j=1}^n (A^t)_{kj} v_j.$$

By the triangular inequality, we have

$$|\lambda v_k| = \left|\sum_{j=1}^n (A^t)_{kj} v_j\right| \leq \sum_{j=1}^n A_{jk}|v_j| \leq |v_k| \sum_{j=1}^n A_{jk} = |v_k|,$$

which implies that $|\lambda| \leq 1$. $\square$

Theorem 2.5. *Suppose that A is a transition matrix with $A_{ij} > 0$ for all i, j, then* $\dim(E_1) = 1$ *and $|\lambda| < 1$ for any eigenvalues $\lambda \neq 1$ of A.*

Proof. Suppose v is an eigenvector of A^t corresponding to λ with $|\lambda| = 1$. It suffices to show that $\lambda = 1$. Since we used

$$|v_k| = |\lambda v_k| = \left|\sum_{j=1}^{n} (A^t)_{kj} v_j\right| \le \sum_{j=1}^{n} A_{jk}|v_j| \le |v_k| \sum_{j=1}^{n} A_{jk} = |v_k|,$$

' the two inequality should be equality. From

$$\sum_{j=1}^{n} A_{jk}|v_j| = |v_k| \sum_{j=1}^{n} A_{jk},$$

we deduce that $v_1 = \cdots = v_n$. From

$$\left|\sum_{j=1}^{n} (A^t)_{kj} v_j\right| = \sum_{j=1}^{n} A_{jk}|v_j|,$$

we see that v_i have the same sign. Thus, we conclude that $v = c(1, 1, \cdots, 1)$ and so $\lambda = 1$. □

Theorem 2.6. *If A is a transition matrix with $A_{ij} > 0$ for all i, j, then*

(i) *the limit $\lim_{n\to\infty} A^n =: L$,*

(ii) *$AL = L = LA$, and*

(iii) *L is a transition matrix such that*

$$L = \begin{pmatrix} | & | & \cdots & | \\ u & u & \cdots & u \\ | & | & \cdots & | \end{pmatrix}$$

where u is the unique probability vector in the eigenspace E_1.

Proof. (i) To prove this, we need the triangularization (or Jordan form) which will be covered later. At this moment, we only prove this when A is diagonalizable. Suppose $A = QDQ^{-1}$ where D is diagonal and Q is invertible. In fact, the previous theorem tells us that

$$D = \mathrm{diag}(1, \lambda_2, \cdots, \lambda_n)$$

with $|\lambda_i| < 1$ for all $i = 2, 3, \cdots, n$. Thus,

$$\lim_{n\to\infty} A^n = \lim_{n\to\infty} QD^nQ^{-1} = \lim_{n\to\infty} Q\,\mathrm{diag}(1, \lambda_2^n, \cdots, \lambda_n^n)Q^{-1} = Q\,\mathrm{diag}(1, 0, \cdots, 0)Q^{-1} =: L$$

(ii) It follows from

$$AL = A \lim_{n\to\infty} A^n = \lim_{n\to\infty} A^{n+1} = L.$$

(iii) Let $u = (1, 1, \cdots, 1)$. Since A^n is a transition matrix (HW), we have $uA^n = u$. It then follows that

$$uL = u \lim_{n\to\infty} A^n = \lim_{n\to\infty} uA^n = \lim_{n\to\infty} u = u,$$

which implies that L is a transition matrix. From Part (ii), we deduce that every column of L is an eigenvector of A associated to 1. Since $\dim(E_1) = 1$ and L is a transition matrix, every column should be the unique probability vector in E_1.

□

3 Inner product spaces

Definition 3.1. Let V be a vector space over $\mathbb{R}$. An inner product on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ such that

(i) $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$ for all $x, y, z \in V$,

(ii) $\langle cx, y \rangle = c \langle x, y \rangle$ for all $x, y \in V$ and $c \in \mathbb{R}$,

(iii) $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in V$, and

(iv) $\langle x, x \rangle > 0$ if $x \neq 0$.

Remark 3.2. Since $\langle 0, 0 \rangle = 0$, one can say that $\langle x, x \rangle \geq 0$ and equality holds if and only if $x = 0$.

Example 3.3. Let $V = \mathbb{R}^n$ and define $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$$

where $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$.

Example 3.4. Let $C([0,1])$ be the set of all continuous functions $f : [0,1] \to \mathbb{R}$. We denote by

$$C([0,1]) = \{f : [0,1] \to \mathbb{R} : f \text{ is continuous.}\}.$$

We define $\langle \cdot, \cdot \rangle : C([0,1]) \times C([0,1]) \to \mathbb{R}$ by

$$\langle f, g \rangle = \int_0^1 f(x) g(x)\, dx$$

for $f, g \in C([0,1])$. Then, one can see that this is an inner product. (The continuity is necessary for the condition (iv).)

Definition 3.5. Let V be a vector space over $\mathbb{C}$. An inner product on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ satisfying (i), (ii), (iv), and

(iii)' $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for all $x, y \in V$,

where the bar denotes the complex conjugate (that is, $\overline{a + bi} = a - bi$ for $a, b \in \mathbb{R}$).

Example 3.6. Let $V = \mathbb{C}^n$. If $n = 2$ and $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ is defined by

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2.$$

Then, this is NOT an inner product. Note that

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i \overline{y_i}$$

is an inner product.

Definition 3.7. Let $A \in \mathcal{M}_{m \times n}(\mathbb{C})$. We define the adjoint of A by

$$A^* = \overline{A^t}.$$

Example 3.8. If $A = \begin{pmatrix} i & 1+2i \\ 2 & 3+4i \end{pmatrix}$, then

$$A^* = \begin{pmatrix} -i & 2 \\ 1-2i & 3-4i \end{pmatrix}.$$

From now on, we consider a vector space V over F where $F = \mathbb{R}$ or $F = \mathbb{C}$.

Example 3.9. Let $A, B \in \mathcal{M}_{n\times n}(F)$. The Frobenius inner product on $\mathcal{M}_{n\times n}(F)$ is defined by

$$\langle A, B\rangle = \operatorname{tr}(AB^*).$$

One can check that this is indeed an inner product. Note that $\langle A, B\rangle = \operatorname{tr}(AB^t)$ if $F = \mathbb{R}$.

Definition 3.10. Let V be a vector space over F. We call V an inner product space over F if it is equipped with a specific inner product $\langle \cdot, \cdot\rangle : V \times V \to F$.

Theorem 3.11. *Let V be an inner product space.*

(i) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$ *for all* $x, y, z \in V$.

(ii) $\langle x, cy\rangle = \overline{c}\,\langle x, y\rangle$ *for all* $x, y \in V$ *and* $c \in F$.

(iii) $\langle x, 0\rangle = \langle 0, x\rangle = 0$ *for all* $x \in V$.

(iv) $\langle x, x\rangle = 0$ *if and only if* $x = 0$.

(v) *If* $\langle x, y\rangle = \langle x, z\rangle$ *for all* $x \in V$, *then* $y = z$.

Proof. Exercise. □

Definition 3.12. Let V be an inner product space over F. We define the norm or length of $x \in V$ by

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

Example 3.13. If $V = \mathbb{C}^n$ is equipped with the standard inner product, then the norm of x is

$$\|x\| = \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}}$$

where $|x_i|^2 = x_i\overline{x_i}$.

Example 3.14. If $V = \mathcal{M}_{n\times n}(F)$ is equipped with the Frobenius inner product, then the norm of a matrix A is

$$\|A\| = \sqrt{\operatorname{tr}(AA^*)}.$$

Theorem 3.15. *Let V be an inner product space over F.*

(i) $\|cx\| = |c|\|x\|$ *for all* $x \in V$ *and* $c \in F$.

(ii) $\|x\| \geq 0$ *for all* $x \in V$ *and equality holds if and only if* $x = 0$.

(iii) *(Cauchy–Schwarz inequality)* $|\langle x, y\rangle| \leq \|x\|\|y\|$ *for all* $x, y \in V$. *Equality holds if and only if* $x = cy$ *for some* $c \in F$.

(iv) *(Triangular inequality)* $\|x + y\| \leq \|x\| + \|y\|$ *for all* $x, y \in V$. *Equality holds if and only if* $x = cy$ *for some* $c \in F$.

Proof. (i) HW.

(ii) HW.

(iii) Let $x, y \in V$ and $c \in F$. If $y = 0$, then it is trivial. Suppose $\|y\| \neq 0$. By Part (ii), we have $\|x - cy\| \geq 0$, which implies

$$\begin{aligned} \|x - cy\|^2 &= \langle x - cy, x - cy \rangle = \langle x, x - cy \rangle - c \langle y, x - cy \rangle \\ &= \langle x, x \rangle - \overline{c} \langle x, y \rangle - c \langle y, x \rangle + |c|^2 \langle y, y \rangle . \end{aligned}$$

Let $c = \langle x, y \rangle / \|y\|^2$, then

$$\begin{aligned} \|x - cy\|^2 &= \langle x, x \rangle - \frac{\overline{\langle x, y \rangle}}{\|y\|^2} \langle x, y \rangle - \frac{\langle x, y \rangle}{\|y\|^2} \langle y, x \rangle + \frac{|\langle x, y \rangle|^2}{\|y\|^2} \\ &= \|x\|^2 - 2\frac{|\langle x, y \rangle|^2}{\|y\|^2} + \frac{|\langle x, y \rangle|^2}{\|y\|^2} \\ &= \frac{\|x\|^2\|y\|^2 - |\langle x, y \rangle|^2}{\|y\|^2} \geq 0. \end{aligned}$$

Thus, we conclude that $\|x\|\|y\| \geq |\langle x, y \rangle|$.

(iv) It follows from Part (iii) that

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \|x\|^2 + \langle x, y \rangle + \langle y, x \rangle + \|y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

□

References

[FIS] Freidberg, Insel, and Spence, *Linear Algebra*, 4th edition, 2002.

[Bee] Beezer, *A First Course in Linear Algebra*, Version 3.5, 2015.

Department of Mathematics, University of Illinois at Urbana-Champaign
E-mail address: daesungk@illinois.edu