# Homework Week #5

Ho Chi Vuong
AI Math Foundations: Abstract Vector Spaces
Center of Talent in AI

April 26, 2020

## Section P

### P4.1

**Consider linear operator** $f : \mathbb{R}^2 \to \mathbb{R}^2$ **with** $f(v_1) = v_1 + 2v_2$ **and** $f(v_2) = 2v_1 + v_2$ **(1). Verify that** $f(v_1 + v_2) = 3(v_1 + v_2)$ **and** $f(v_1 - v_2) = -(v_1 - v_2)$ **(2)**

1. $f$ **stretches vectors in the** $v_1 + v_2$ **direction by a factor of 3.**

2. $f$ **negates vectors in the** $v_1 - v_2$ **direction.**

**Express** $f$ **using basis** $B_v = \{v_1, v_2\}$**, then** $B_e = \{e_1 = v_1 + v_2, e_2 = v_1 - v_2\}$**:**

1. $f(av_1 + bv_2) = (a + 2b)v_1 + (2a + b)v_2$**: coefficients** $a$ **&** $b$ **are "mixed up".**

2. $f(ae_1 + be_2) = 3ae_1 - be_2$**: coefficients** $a$ **and** $b$ **are simply scaled.**

**Clearly, Eq (2) is a better representation than Eq (1) to understand the geometry of** $f$**. What is the matrix representation of** $f$ **in** $B_v$**? in** $B_e$**?**

1. Verify that $f(v_1+v_2) = 3(v_1+v_2)$ and $f(v_1-v_2) = -(v_1-v_2)$
   Using the linearity of the operator, we have the following:

$$f(v_1+v_2) = f(v_1) + f(v_2) = v_1 + 2v_2 + 2v_1 + v_2 = 3(v_1+v_2)$$

   and

$$f(v_1-v_2) = f(v_1) - f(v-2) = v_1 + 2v_2 - 2v_1 - v_2 = -(v_1-v_2)$$

## P4.2

**$A$ is a Markov matrix, i.e., each column $\mathbf{a}_i$ is a probability vector. Show that if $p_0$ is a probability $n$-vector then $Ap_0$ is also a probability vector.**
Suppose that $A$ is a $m \times n$ matrix. We have that:

$$Ap_0 = \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \\ A_{m1} & ldots & A_{mn} \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix} = \begin{bmatrix} p_1a_{11} + \ldots + p_na_{1n} \\ \vdots \\ p_1a_{m1} + \ldots + p_na_{mn} \end{bmatrix}$$

Taking the sum of the resulting matrix, we have:

$$\begin{aligned} &p_1a_{11} + \ldots + p_na_{1n} + \ldots + p_1a_{m1} + \ldots + p_na_{mn} \\ =&p_1(a_{11} + \ldots + a_{m1}) + \ldots + p_n(a_{11} + \ldots + a_{mn}) \end{aligned}$$

As each column $\mathbf{a}_n$ is a probability vector $\Leftrightarrow a_{1n} + \ldots + a_{mn} = 1$ and $p_0$ is a probability matrix, we then continue the above equation:

$$\begin{aligned} =&p_1 + \ldots + p_n \\ =&1 \end{aligned}$$

Thus making $Ap_0$ also a probability matrix.

## P4.4

**Prove again if $\beta = \{v_1, \ldots, v_n\}$ is an eigenbasis of $A_{n\times n}$ with eigenvalues $\lambda_i$'s, then $A$ is diagonalizable: $A = QDQ^{-1}$ with $D = \textsf{diag}(\lambda_1, \ldots, \lambda_n)$ and $Q$ a square matrix composed of $v_i$'s as columns.**
As $\beta$ is an eigenbasis of $A$, then we have:

$$Av_j = \lambda_j v_j$$

which means:

$$\begin{aligned}
A\begin{bmatrix} v_1 & \dots & v_j \end{bmatrix} &= \begin{bmatrix} \lambda_1 v_1 & \dots & \lambda_n v_n \end{bmatrix} \\
&= diag(\lambda_1, \dots, \lambda_n) \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \\
&= \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} diag(\lambda_1, \dots, \lambda_n) \\
\Leftrightarrow A &= \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} diag(\lambda_1, \dots, \lambda_n) \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}^{-1}
\end{aligned}$$

Replacing $\begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} = Q$ and $D = diag(\lambda_1, \dots, \lambda_n)$, we have:

$$A = QDQ^{-1}$$

as desired.

## P3.4

**Polarization identity:** $\langle v, w \rangle = \frac{1}{2}\big(\|v+w\|^2 - \|v\|^2 - \|w\|^2\big)$
We have the right-hand side of the equation:

$$\begin{aligned}
\frac{1}{2}\big(\|v+w\|^2 - \|v\|^2 - \|w\|^2\big) &= \frac{1}{2}\big(\langle v+w, v+w \rangle - \langle v, v \rangle - \langle w, w \rangle\big) \\
&= \frac{1}{2}\big(\langle v, v \rangle + \langle v, w \rangle + \langle w, v \rangle + \langle w, w \rangle - \langle v, v \rangle - \langle w, w \rangle\big) \\
&= \frac{1}{2}\big(\langle v, w \rangle + \langle w, v \rangle\big) \\
&= \frac{1}{2}\big(\|v\|\|w\|cos(v,w) + \|w\|\|v\|cos(w,v)\big) \\
&= \frac{1}{2}\big(2\|w\|\|v\|cos(v,w)\big) \\
&= \langle v, w \rangle
\end{aligned}$$

which is equal to the left-hand side. We thus prove the Polarization identity.

## P4.5

**Summarize the *problem description* and *solution* to find $F_n$ in Fibonacci's rabbits $F_0 = 0, F_1 = 1, \dots, F_n = F_{n-1} + F_{n-2}$.**

Distance from the point $w$ to the plane $H_n$ is:

$$D(w, H_n) = \frac{aw_x + bw_y + cw_z - (ax_0 + by_0 + cz_0)}{\sqrt{a^2 + b^2 + c^2}}$$

Assume for contradiction that there exists a point $w_{p1} \in H_n$ beside $w_p$ (the orthogonal projection of $w$ on $H_n$) whose length $ww_{p1} < ww_p$. According to the Pythagorean theorem, we have:

$$ww_{p1}^2 = ww_p^2 + w_{p1}w_p^2$$
$$\Rightarrow ww_{p1} > ww_p$$

which contradicts the previous statement. Thus, $ww_p$ is the smallest distance to any point in $H_n$.

## P4.6

**Show that $\lambda$ is an eigenvalue of a square $n \times n$ matrix $A$ iff $\mathsf{det}(A - \lambda I_n) = 0$.**
We have $\lambda$ is an eigenvalue of a $A_{n\times n}$ when, for any eigenvector $v$:

$$Av = \lambda v$$
$$\Leftrightarrow Av - \lambda I v = 0$$
$$\Leftrightarrow (A - \lambda I)v = 0$$

Geometrically, this transformation transforms $v$ into a 0 vector, and this is only possible when $\mathsf{det}(A - \lambda I) = 0$. We thus have as desired.

## P4.7-4.9

**Given a *symmetric* positive definite matrix $A > 0$, i.e., $x^\top Ax > 0 \ \forall x \neq 0_n$.**

## P4.7

**Show that all its eigenvalues are positive, $\lambda_i > 0 \ \forall i$.**
Suppose that $v$ is an eigenvector of $A_{n\times n}$ with eigenvalue $\lambda$, then we have:

$$Av = \lambda v$$
$$\Leftrightarrow v^\mathsf{T} Av = v^\mathsf{T} \lambda v$$
$$\Rightarrow v^\mathsf{T} \lambda v > 0$$
$$\Rightarrow \lambda(v_1^2 + \ldots + v_n^2) > 0$$

and as $(v_1^2 + \ldots + v_n^2) > 0$, $\forall v_j \in R^n$,

$$\Rightarrow \lambda > 0$$

## P3.8

**Show that the generalized dot product $\langle \mathbf{x}, \mathbf{y} \rangle_M = \mathbf{x}^\top M \mathbf{y}$ with $M \in \mathbb{R}^{n\times n}$ a symmetric positive definite matrix satisfies all 3 requirements of a proper inner product.**

Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ ,$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ and $M = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_n \end{bmatrix}$ with $\mathbf{m}_i$, $i = 1, 2, \ldots, n$, having $n$ elements. For any $\mathbf{x}, \mathbf{x}'$ in the same inner product space, $c \in R$, we have the following 3 requirements of an inner product:

1. $\langle \mathbf{x} + \mathbf{x}', y \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}', \mathbf{y} \rangle$ and $\langle c\mathbf{x}, \mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle$
   We have the following:

$$\begin{aligned}
\langle \mathbf{x} + \mathbf{x}', \mathbf{y} \rangle &= \begin{bmatrix} x_1 + x_1' & x_2 + x_2' & \ldots & x_n + x_n' \end{bmatrix} M\mathbf{y} \\
&= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} M\mathbf{y} + \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix} M\mathbf{y} \\
&= \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}', \mathbf{y} \rangle
\end{aligned}$$

   and

$$\begin{aligned}
\langle c\mathbf{x}, y \rangle &= \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix} My \\
&= c \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} My \\
&= c\langle \mathbf{x}, \mathbf{y} \rangle
\end{aligned}$$

2. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$

$$
\begin{aligned}
\langle \mathbf{x}, \mathbf{y} \rangle &= \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_n \end{bmatrix} \mathbf{y} \\
&= x_1\mathbf{m}_1\mathbf{y} + x_2\mathbf{m}_2\mathbf{y} + \dots + x_n\mathbf{m}_n\mathbf{y} \\
&= y_1\mathbf{m}_1\mathbf{x} + y_2\mathbf{m}_2\mathbf{x} + \dots + y_n\mathbf{m}_n\mathbf{x} \\
&= \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_n \end{bmatrix} \mathbf{x} \\
&= \langle \mathbf{y}, \mathbf{x} \rangle
\end{aligned}
$$

3. $\langle \mathbf{x}, \mathbf{x} \rangle > 0$
   Since M is a positive definite matrix, it is obvious that:

$$\langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^{\mathsf{T}} M \mathbf{x} > 0$$

With the 3 requirements proved, we have the above.

## P3.9

**Prove that the null space and the row space of a matrix are orthogonal, i.e. every vector in null space is orthogonal to every vector in row space (zero dot product).**

Assume we have a matrix $A = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$, its null space $N(A)$ which contains vectors $x$ satisfying $Ax = 0$, and its row space $R = span(v_1, \dots, v_n)$. We have $Ax = 0 \Leftrightarrow v_1 x = 0, \dots, v_n x = 0$. Therfore, for $r \in R$, we have the following:

$$rx = span(v_1, \dots, v_n)x = a_1 v_1 x + \dots + a_n v_n x = 0$$

From here we have proven that $rx = 0 \forall r \in R$, which means every $r$ is ortogonal with $x$, which proves what we have to prove.

Assume we have a matrix $A = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$, its null space $N(A)$ which contains vectors $x$ satisfying $Ax = 0$, and its row space $R = span(v_1, \ldots, v_n)$. We have $Ax = 0 \Leftrightarrow v_1 x = 0, \ldots, v_n x = 0$. Therfore, for $r \in R$, we have the following:

$$rx = span(v_1, \ldots, v_n)x = a_1 v_1 x + \ldots + a_n v_n x = 0$$

From here we have proven that $rx = 0$, $\forall r \in R$, which means every $r$ is ortogonal with $x$, which proves what we have to prove.

### P3.10

## Section E

### E4.1

**Show that $\forall i : |\lambda_i| \leq 1$ & $A$ always has an eigenvalue $\lambda_1 = 1$ (single one if $A$ is positive definite**

Suppose that $A$ is a $n \times n$ matrix, then we have that:

$$Av = \lambda v$$

$$\Rightarrow A \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} a_{11} + \ldots + a_{1n} \\ \ldots \\ a_{n1} + \ldots + a_{nn} \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Then $\exists \lambda = 1$.

## E4.3

**Prove again the **spectral theorem** for matrix representation of linear operator $T$ in basis $\beta$: $[T]_\beta^\beta = \textbf{diag}(\lambda_1, \ldots, \lambda_n) \Leftrightarrow \beta$ is an eigenbasis of $T$ with eigenvalues $\lambda_i$'s.**

We first prove that $\beta$ is an eigenbasis of $T$ with eigenvalues $\lambda_i$'s $\Rightarrow$ $\beta$: $[T]_\beta^\beta = \mathsf{diag}(\lambda_1, \ldots, \lambda_n)$. Supppose that $\beta = \{v_1, \ldots, v_n\}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. We have that $Tv_j = \lambda_j v_j$, so the coordinate matrix $[Tv_j]^\beta$ is $\begin{bmatrix} 0 \\ \vdots \\ \lambda_j \\ \vdots \\ 0 \end{bmatrix}$ with $\lambda_j$ at the $j$th-index. Putting all the $[Tv_j]^\beta$ together, we have $[Tv_j]_\beta^\beta$ is $\begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix} = \mathsf{diag}(\lambda_1, \ldots, \lambda_n)$.

We now prove that $\beta$: $[T]_\beta^\beta = \mathsf{diag}(\lambda_1, \ldots, \lambda_n) \Rightarrow \beta$ is an eigenbasis of $T$ with eigenvalues $\lambda_i$'s. Similarly, for each $j$th column in $[T]_\beta^\beta = [Tv_j]_\beta^\beta$, we have the coordinate matrix $\begin{bmatrix} 0 \\ \vdots \\ \lambda_j \\ \vdots \\ 0 \end{bmatrix}$, which corresponds to the vector $\lambda_j v_j$. We thus have $[Tv_j]^\beta = [\lambda_j v_j]^\beta \Leftrightarrow Tv_j = \lambda_j v_j$, making each $v_j$ an eigenvector, and $\beta = \{v_1, \ldots, v_n\}$ an eigenbasis.

## P4.4

**Show that if $A_{n\times n}$ has $n$ distinct eigenvalues $\Rightarrow A$ is diagonalizable.** Suppose that $A$ has $n$ distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ which corresponds to the eigenvectors $v_1, \ldots, v_n$. As $lambda_1, \ldots, \lambda_n$ are all distinct, $v_1, \ldots, v_n$ are linearly independent (as proven in Terence Tao's note), and they thus form an eigenbasis of $A$. We can thus express $A$ as a diagonal matrix $\mathsf{diag}(\lambda_1, \ldots, \lambda_n)$ corresponding to $\{v_1, \ldots, v_n\} \Rightarrow A$ is diagonalizable.

## P4.7

**Prove that translation does not change volume.**

Suppose that we have a translation matrix T in 3D $\begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Then the determinant of this matrix is

$$\mathsf{det}(\mathsf{T}) = 1\mathsf{det}(\begin{bmatrix} 1 & 0 & t_y \\ 0 & 1 & t_z \\ 0 & 0 & 1 \end{bmatrix}) + t_x\mathsf{det}(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}) = 1.1 + 0 = 1$$

Using the induction hypothesis, for a translation matrix T in n-dimension, $\begin{bmatrix} 1 & 0 & \ldots & t_1 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_n \\ 0 & 0 & 0 & 1 \end{bmatrix}$, its determinant is

$$\mathsf{det}(T) = 1(1(\ldots(1\mathsf{det}\begin{bmatrix} 1 & 0 & t_y \\ 0 & 1 & t_z \\ 0 & 0 & 1 \end{bmatrix} + t_{n-2}\mathsf{det}(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix})) + t_{n-3}(0 + t_{n-2}\mathsf{det}(\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix})\ldots)\ldots)$$

$$+t_1(0 + t_2(\ldots(0 + t_{n-2}\mathsf{det}\begin{bmatrix} 0 & 0 & t_{n-1} \\ 0 & 0 & t_n \\ 0 & 0 & 1 \end{bmatrix})))$$

$$= 1 \times 1 \times \ldots \times (((1+0)+0)+0)\ldots) + t_1 \times t_2 \times \ldots t_{n-2} \times 0 = 1$$

With the determinant $= 1$, by definition the volume does not change after the translation.