Homework Week #4

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

April 19, 2020

Section P

P3.1

Cauchy-Schwarz inequality: $|\langle v, w\rangle| \leq \|v\|\|w\|$
Beside the solution in Terence Tao's note, we can prove it intuitively as follows. We have the following:

$$|\langle v, w\rangle| \leq \|v\|\|w\|$$

$$\Leftrightarrow (\langle v, w\rangle)^2 \leq \|v\|^2\|w\|^2$$

$$\Leftrightarrow (\|v\|\|w\|cos(v, w))^2 \leq \|v\|^2\|w\|^2$$

$$\Leftrightarrow \|v\|^2\|w\|^2cos(v, w)^2 \leq \|v\|^2\|w\|^2$$

As $-1 \leq cos(v, w) \leq 1$, we have proved the above inequality.

P3.2

Triangle inequality: $\|v\| - \|w\| \leq \|v + w\| \leq \|v\| + \|w\|$
We have the following:

$$\|v\| - \|w\| \leq \|v + w\| \leq \|v\| + \|w\|$$

$$\Leftrightarrow \sqrt{\langle v, v\rangle} - \sqrt{\langle w, w\rangle} \leq \sqrt{\langle v + w, v + w\rangle} \leq \sqrt{\langle v, v\rangle} + \sqrt{\langle w, w\rangle}$$

$$\Leftrightarrow \langle v, v\rangle - 2\sqrt{\langle v, v\rangle\langle w, w\rangle} + \sqrt{\langle w, w\rangle} \leq \langle v, v+w\rangle + \langle w, v+w\rangle \leq \langle v, v\rangle + 2\sqrt{\langle v, v\rangle\langle w, w\rangle} + \sqrt{\langle w, w\rangle}$$

$$\Leftrightarrow \langle v,v\rangle - 2\sqrt{\langle v,v\rangle\langle w,w\rangle} + \langle w,w\rangle \leq \langle v,v\rangle + \langle v,w\rangle + \langle w,v\rangle + \langle w,w\rangle \leq \langle v,v\rangle + 2\sqrt{\langle v,v\rangle\langle w,w\rangle} + \langle w,w\rangle$$

$$\Leftrightarrow -2\sqrt{\langle v,v\rangle\langle w,w\rangle} \leq \langle v,w\rangle + \langle w,v\rangle \leq 2\sqrt{\langle v,v\rangle\langle w,w\rangle}$$

According to Cauchy-Schwarz inequality, we have:

$$-\|v\|\|w\| \leq \langle v,w\rangle \leq \|v\|\|w\|$$
$$-\|w\|\|v\| \leq \langle w,v\rangle \leq \|w\|\|v\|$$

$$\Rightarrow -\|w\|\|v\| - \|w\|\|v\| = -2\|v\|\|w\| \leq \langle v,w\rangle + \langle w,v\rangle \leq \|v\|\|w\| + \|w\|\|v\| = 2\|v\|\|w\|$$

Therefore, we can the above inequality is true, and thus the Triangle Inequality.

P3.3

Parallellogram law: $\|v+w\|^2 + \|v-w\|^2 = 2(\|v\|^2 + \|w\|^2)$
We have the left-hand side of the equation:

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

which is equal to the right-hand side. We thus prove the Parallellogram law.

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.

P3.5

Find the distance from a point $\mathbf{w} = (w_x, w_y, w_z)^\top$ to the plane H_n normal to $\mathbf{n} = (a, b, c)^\top$ and translated from the origin by vector $\mathbf{p} = (x_0, y_0, z_0)^\top$. Show that it is the smallest distance to any point in H_n.

The equation of the plane H_n is:

$$\begin{aligned} H_n &= a(x - x_0) \\ &= a(x - x_0) + b(y - y_0) + c(z - z_0) \\ &= ax + by + cz - (ax_0 + by_0 + cz_0) \end{aligned}$$

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.

P3.6

Prove that $w_p = \mathsf{proj}_{V_k}(w)$ is the vector in subspace V_k closest to w using 2 approaches:

1. **expand $\|u - v\|^2 = \langle u - v, u - v \rangle$.**
2. **using Pythagorean theorem**

In order to prove the above, we have to prove, for any $w_{p1} \in V$, $\|w - w_{p1}\| > \|w - w_p\|$.

1. expand $\|u-v\|^2 = \langle u-v, u-v \rangle$.

2. using Pythagorean theorem
Let $a = w - w_p$, $b = w_p - w_{p1} \Rightarrow b + a = w - w_{p1}$. Since $a \perp b \Leftrightarrow \langle a, b \rangle = \langle b, a \rangle = 0$, we have the following:

$$\begin{aligned}\|w - w_{p1}\|^2 = \|b + a\|^2 &= \langle b + a, b + a \rangle \\ &= \langle b, b \rangle + \langle b, a \rangle + \langle a, b \rangle + \langle a, a \rangle \\ &= \|b\|^2 + 0 + 0 + \|a\|^2 > \|a\|^2 = \|w - w_p\|^2 \\ &\Rightarrow \|w - w_{p1}\| > \|w - w_p\|\end{aligned}$$

We have thus proved the above using Pythagorean theorem.

P3.7

Least squares solution of $A\mathbf{x} = \mathbf{b}$ is $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|$. By column view of matrix multiplication, $A\hat{\mathbf{x}}$ is the orthogonal projection of b onto the column space $\mathsf{Col}(A)$ of matrix A. Thus we must have $(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathsf{Col}(A)$. Write down matrix form for this and solve for $\hat{\mathbf{x}}$, given that A is full rank. Show that the orthogonal projection operator of b onto $\mathsf{Col}(A)$ is: $P_{\mathsf{Col}(A)} = A(A^\top A)^{-1}A^\top$.

Assume that $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}$ are the columns of A, we have $(\mathbf{b} - A\hat{\mathbf{x}}) \perp \mathsf{Col}(A)$ in matrix form is:

$$[\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - [\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}] \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \end{bmatrix}$$

We have that $\mathbf{b} = \mathbf{b}_{\mathsf{Col}(A)} + \mathbf{b}_{\mathsf{Col}(A)^\perp}$, and $\mathbf{b}_{\mathsf{Col}(A)} = A\hat{\mathbf{x}}$, so $\mathbf{b} - \mathbf{b}_{\mathsf{Col}(A)} = \mathbf{b} - A\hat{\mathbf{x}}$, which also equals to zero, because $\mathbf{b} - \mathbf{b}_{\mathsf{Col}(A)} = \mathbf{b}_{\mathsf{Col}(A)^\perp} \in \mathsf{Col}(A)^\perp = N(\mathsf{Col}(A))$. We then have:

$$\begin{aligned} 0 &= A^\mathsf{T}(\mathbf{b} - A\hat{\mathbf{x}}) = A^\mathsf{T} b - A^\mathsf{T} A\hat{\mathbf{x}} \\ &\Leftrightarrow A^\mathsf{T}\mathbf{b} = A^\mathsf{T} A\hat{\mathbf{x}} \\ &\Leftrightarrow \hat{\mathbf{x}} = (A^\mathsf{T} A)^{-1} A^\mathsf{T}\mathbf{b} \end{aligned}$$

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

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

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 $R = span(v_1, \ldots, v_n)$. 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.

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 $Ax = 0 \Leftrightarrow v_1x = 0, \ldots, v_nx = 0$. Therfore, for $r \in R$, we have the following:

$$rx = span(v_1, \ldots, v_n)x = a_1v_1x + \ldots + a_nv_nx = 0$$

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

E3.1

Let V be an n-dimensional inner product space with pairwise orthogonal subspaces

$$W_1, \ldots, W_m\,,$$

where $\sum_{i=1}^{m} \dim(W_i) = n$. Prove that every vector $v \in V$ can be represented uniquely as

$$v = w_1 + \cdots + w_m\,,$$

where $w_i \in W_i$ for $i = 1, \ldots, m$, i.e.,

$$V = W_1 \oplus \cdots \oplus W_m\,.$$

Assume that V has basis vectors $\{v_1, v_2, \ldots, v_n\}$, and W_1 has bases $\{v_1, v_2, \ldots, v_i\}$, W_2 has bases $\{v_{i+1}, v_{i+2}, \ldots v_j\}$, $\ldots$, W_m has bases $\{v_{k+1}, v_{k+2}, \ldots, v_n\}$. We have any $v \in V$, $w_1 \in W_1$, $w_2 \in W_2$,...,$w_m \in W_m$, for any $a_1, a_2, \ldots a_i, a_{i+1}, a_{i+2}$, $\ldots, a_j, \ldots, a_{k+1}, a_{k+2}, \ldots, a_n$ can be expressed as follow:

$$\begin{aligned}
v &= a_1v_1 + a_2v_2 + \ldots + a_nv_n \\
w_1 &= a_1v_1 + a_2v_2 + \ldots + a_iv_i \\
w_2 &= a_{i+1}v_{i+1} + a_{i+2}v_{i+2} + \ldots + a_jv_j \\
&\vdots \\
w_m &= a_{k+1}v_{k+1} + a_{k+2}v_{k+2} + \ldots + a_nv_n
\end{aligned}$$

We then have:

$$\begin{aligned}
w_1 + w_2 + \ldots + w_m =& a_1v_1 + a_2v_2 + \ldots + a_iv_i \\
&+ a_{i+1}v_{i+1} + a_{i+2}v_{i+2} + \ldots + a_jv_j \\
&\ldots \\
&+ a_{k+1}v_{k+1} + a_{k+2}v_{k+2} + \ldots + a_nv_n \\
&= v
\end{aligned}$$

since $\sum_{i=1}^{m} \dim(W_i) = n$. Therefore $v = w_1 + w_2 + \ldots + w_m$ uniquely, for an unique set of $a_1, a_2, \ldots, a_n$.

E3.2

Prove again for 2D geometry that $\cos\theta = \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$ using only cosine (and sine) laws.

$$\begin{aligned}
cos(\theta) = cos(\beta - \alpha) &= cos(\beta)cos(\alpha) + sin(\beta)sin(\alpha) \\
&= \frac{b_1}{\|\mathbf{b}\|}\frac{a_1}{\|\mathbf{a}\|} + \frac{b_2}{\|\mathbf{b}\|}\frac{a_2}{\|\mathbf{a}\|} \\
&= \frac{b_1a_1}{\|\mathbf{b}\|\|\mathbf{a}\|} + \frac{b_2a_2}{\|\mathbf{b}\|\|\mathbf{a}\|} \\
&= \frac{b_1a_1 + b_2a_2}{\|\mathbf{b}\|\|\mathbf{a}\|} \\
&= \frac{\begin{bmatrix} a_1 & a_2 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}{\|\mathbf{b}\|\|\mathbf{a}\|} \\
&= \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}
\end{aligned}$$

E3.3

Given an orthonormal basis $\mathcal{U} = (u_1, \ldots, u_n)$ of $\mathbb{R}^n$. Let U be the matrix whose distinct columns are the basis vectors in $\mathcal{U}$. Prove again U is a transformation that does not change distance, angle, size of the objects being transformed: geometrically it is a rotation or reflection!

Let A be the object being transformed, and $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$. Since U is a matrix consisting of orthonormal vector $\Leftrightarrow U^\top U = UU^\top = I_n$, then we have the following:

$$\langle U\mathbf{a}_i, U\mathbf{a}_j\rangle = U\langle \mathbf{a}_i, U\mathbf{a}_j\rangle = \langle \mathbf{a}_i; U^\top U\mathbf{a}_j\rangle = \langle \mathbf{a}_i, \mathbf{a}_j\rangle$$

for any $i, j = 1, 2, \ldots, n$. We can conclude that the distance does not change after transforming.

We can prove that the angle and the size does not change by the following:

$$\|U\mathbf{a}_i\|^2 = \langle U\mathbf{a}_i, U\mathbf{a}_i\rangle = \langle \mathbf{a}_i, \mathbf{a}_i\rangle = \|\mathbf{a}_i\|^2$$

$$cos(U\mathbf{a}_i, U\mathbf{a}_j) = \frac{\langle U\mathbf{a}_i, U\mathbf{a}_j\rangle}{\|U\mathbf{a}_i\|\|U\mathbf{a}_j\|} = \frac{\langle \mathbf{a}_i, \mathbf{a}_j\rangle}{\|\mathbf{a}_i\|\|\mathbf{a}_j\|} = \cos(\mathbf{a}_i, \mathbf{a}_j)$$