DS-GA 1014 - Homework 5

Eric Niblock

October 5th, 2020

1. (2 points). Let S be a subspace of $\mathbb{R}^n$ and let P_S be the matrix of the orthogonal projection onto S. Let $M = Id_n - 2P_S$.

(a) Show that the matrix M is orthogonal.

First, we have the following proposition,

Let $A \in \mathbb{R}^{n \times n}$. Then if A is an orthogonal matrix, this is equivalent to saying $AA^T = Id_n = A^T A$ [Prop. 1].

So to show that matrix M is orthogonal, all we must show is that $MM^T = Id_n$. This follows readily,

$$MM^T = (Id_n - 2P_S)(Id_n - 2P_S)^T = (Id_n - 2P_S)(Id_n^T - 2P_S^T) \tag{1}$$

Now, we must employ two properties. The first is that, for any projection matrix P, we have that $P = P^T$ (and it goes without saying that $Id_n = Id_n^T$). Secondly, we have that for any projection matrix $P = P^2$ (and it goes without saying that $Id_n = Id_n^2$). Then, it follows from the end of (1) that,

$$\begin{aligned} MM^T &= (Id_n - 2P_S)(Id_n^T - 2P_S^T) = (Id_n - 2P_S)(Id_n - 2P_S) \\ &= Id_n^2 - 2Id_nP_S - 2P_SId_n + 4P_S^2 \\ &= Id_n - 4P_S + 4P_S \\ &= Id_n \end{aligned} \tag{2}$$

This suffices to show that M is an orthogonal matrix by *Proposition 1*, since $MM^T = Id_n$.

(b) Show that if $\lambda \in \mathbb{R}$ is an eigenvalue of M, then $\lambda = 1$ or $\lambda = -1$.

We know that M is an orthogonal matrix, and hence preserves the length of any vector is acts upon, since,

$$||M\overrightarrow{\mathbf{v}}||^2 = \langle M\overrightarrow{\mathbf{v}}, M\overrightarrow{\mathbf{v}}\rangle = \overrightarrow{\mathbf{v}}^T M^T M\overrightarrow{\mathbf{v}} = \overrightarrow{\mathbf{v}}^T\overrightarrow{\mathbf{v}} = \langle\overrightarrow{\mathbf{v}}, \overrightarrow{\mathbf{v}}\rangle = ||\overrightarrow{\mathbf{v}}||^2 \tag{3}$$

Now, if we take $\overrightarrow{\mathbf{v}} \in \mathbb{R}^{\mathrm{n}}$ to be some eigenvector of A, we also have,

$$M\overrightarrow{\mathbf{v}} = \lambda\overrightarrow{\mathbf{v}} \tag{4}$$

$$\langle M\overrightarrow{\mathbf{v}}, M\overrightarrow{\mathbf{v}}\rangle = \langle\lambda\overrightarrow{\mathbf{v}}, \lambda\overrightarrow{\mathbf{v}}\rangle = \lambda^2\langle\overrightarrow{\mathbf{v}}, \overrightarrow{\mathbf{v}}\rangle = \lambda^2||\overrightarrow{\mathbf{v}}||^2 \tag{5}$$

This implies that,

$$||\overrightarrow{\mathbf{v}}||^2 = \lambda^2||\overrightarrow{\mathbf{v}}||^2 \tag{6}$$

$$\lambda = \pm 1 \tag{7}$$

2. (2 points). Let $\vec{\mathbf{v}} \in \mathbb{R}^n$ be a non-zero vector. What are the eigenvalues of the $n \times n$ matrix

$$M = \vec{\mathbf{v}}\vec{\mathbf{v}}^T$$

and their multiplicities? (In the expression $\vec{\mathbf{v}}\vec{\mathbf{v}}^T$ we see $\vec{\mathbf{v}}$ as a matrix with 1 column and n rows).

We can write the matrix M with columns that are combinations of the rows of $\vec{\mathbf{v}}$ as follows,

$$M = \vec{\mathbf{v}}\vec{\mathbf{v}}^T = \begin{bmatrix} v_1^T \vec{\mathbf{v}} & \Bigg| & ... & \Bigg| & v_n^T \vec{\mathbf{v}} \end{bmatrix} \tag{8}$$

So it is evident that the columns of M are linearly dependent, since each is just a constant multiple of $\vec{\mathbf{v}}$. Since $\vec{\mathbf{v}} \neq \vec{\mathbf{0}}$, we have that $rank(M) = 1$.

Now, we can very easily find one eigenvalue since,

$$M\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}} \tag{9}$$

$$(\vec{\mathbf{v}}\vec{\mathbf{v}}^T)\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}} \tag{10}$$

$$\vec{\mathbf{v}}(\vec{\mathbf{v}}^T\vec{\mathbf{v}}) = \lambda\vec{\mathbf{v}} \tag{11}$$

So, it follows that the eigenvalue associated with the eigenvector $\vec{\mathbf{v}}$ is given by,

$$\lambda = \vec{\mathbf{v}}^T\vec{\mathbf{v}} \tag{12}$$

Now, an eigenvalue of 0 implies that,

$$M\vec{\mathbf{u}} = 0 \cdot \vec{\mathbf{u}} = \vec{\mathbf{0}} \tag{13}$$

Which implies that any vector $\vec{\mathbf{u}} \in \mathcal{N}(M)$ is actually an eigenvector associated with an eigenvalue of zero. We now take advantage of the Rank-Nullity Theorem, *If A is some $n \times m$ matrix, then*

$$rank(A) + dim(Ker(A)) = m \tag{14}$$

[Thm. 1]

So, since we already know $rank(M) = 1$, it follows that $dim(\mathcal{N}(M)) = n - 1$. This implies there are $n-1$ eigenvectors associated with $\lambda = 0$, meaning it has a multiplicity of $n - 1$. Therefore we have $\lambda = \overrightarrow{\mathbf{v}}^T \overrightarrow{\mathbf{v}}$, with a multiplicity of one, and $\lambda = 0$ with a multiplicity of $n-1$. There can be no more eigenvalues, since the sum of multiplicities must be less than or equal to n.

3. (2 points). Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Show that if $\vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \in \mathbb{R}^n$ are two eigenvectors of A associated to distinct eigenvalues $\lambda_1 \neq \lambda_2$ ($A\vec{\mathbf{v_1}} = \lambda_1 \vec{\mathbf{v_1}}$ and $A\vec{\mathbf{v_2}} = \lambda_2 \vec{\mathbf{v_2}}$), then $\vec{\mathbf{v_1}} \perp \vec{\mathbf{v_2}}$.

Equivalently, we must show that $\langle \vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle = 0$ since $\langle \vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle = 0 \iff \vec{\mathbf{v_1}} \perp \vec{\mathbf{v_2}}$. We begin with an expression to show a useful equality. Within, we exploit the fact that A is symmetric, such that $A = A^T$,

$$\begin{aligned} \lambda_1 \langle \vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle &= \langle \lambda_1 \vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle = \langle A\vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle = (A\vec{\mathbf{v_1}})^T \cdot \vec{\mathbf{v_2}} \\ &= \vec{\mathbf{v_1}}^T A^T \cdot \vec{\mathbf{v_2}} = \vec{\mathbf{v_1}}^T (A \cdot \vec{\mathbf{v_2}}) = \langle \vec{\mathbf{v_1}}, A\vec{\mathbf{v_2}} \rangle \\ &= \langle \vec{\mathbf{v_1}}, \lambda_2 \vec{\mathbf{v_2}} \rangle = \lambda_2 \langle \vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle \end{aligned} \tag{15}$$

So, we are left with the resulting equality,

$$\lambda_1 \langle \vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle = \lambda_2 \langle \vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle \tag{16}$$

$$(\lambda_1 - \lambda_2) \langle \vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle = 0 \tag{17}$$

At which point we know either $\lambda_1 - \lambda_2 = 0$ or $\langle \vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle = 0$. However, from the statement of the problem we know that $\lambda_1 \neq \lambda_2$, so it follows that,

$$\langle \vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \rangle = 0 \iff \vec{\mathbf{v_1}} \perp \vec{\mathbf{v_2}} \tag{18}$$

And therefore, if $\vec{\mathbf{v_1}}, \vec{\mathbf{v_2}} \in \mathbb{R}^n$ are two eigenvectors of a symmetric matrix A associated to distinct eigenvalues $\lambda_1 \neq \lambda_2$, then $\vec{\mathbf{v_1}} \perp \vec{\mathbf{v_2}}$.

4. (4 points). Let $A \in \mathbb{R}^{n \times n}$. We assume that there exists a basis $(\overrightarrow{\mathbf{v}_1}, ..., \overrightarrow{\mathbf{v}_n})$ of $\mathbb{R}^n$ consisting of eigenvectors of A:

$$A\overrightarrow{\mathbf{v}_i} = \lambda_i \overrightarrow{\mathbf{v}_i}$$

for all $i \in \{1, ..., n\}$. We assume that

$$\lambda_1 > |\lambda_i| \qquad \textbf{for all } i \in \{2, ..., n\}$$

We consider the following algorithm:

- **Initialize $x_0 \in \mathbb{R}^n$.**
- **Perform the updates: $x_{t+1} = \frac{Ax_t}{||Ax_t||}$**

(a) Show that for all $t \geq 1$,

$$x_t = \frac{A^t x_0}{||A^t x_0||}$$

We show this via induction (vector indicators removed for convenience). First, we prove the base case. For x_1, we have,

$$x_1 = \frac{A^1 x_0}{||A^1 x_0||} = \frac{Ax_0}{||Ax_0||} \tag{19}$$

And by the update method, with $t = 0$, we find,

$$x_{0+1} = x_1 = \frac{Ax_0}{||Ax_0||} \tag{20}$$

Thus, we have shown the base case. Now, for the inductive step, we find that the update method is given by,

$$x_{t+1} = \frac{Ax_t}{||Ax_t||} = \frac{A\frac{A^t x_0}{||A^t x_0||}}{||A\frac{A^t x_0}{||A^t x_0||}||} = \frac{\frac{1}{||A^t x_0||}}{\frac{1}{||A^t x_0||}} \frac{A^{t+1} x_0}{||A^{t+1} x_0||} = \frac{A^{t+1} x_0}{||A^{t+1} x_0||} \tag{21}$$

This is the result we hoped for, because it is equivalent to,

$$x_{(t+1)} = \frac{A^{(t+1)} x_0}{||A^{(t+1)} x_0||} \tag{22}$$

Thus we have shown the inductive step, because we guessed the method worked for every x_t, and proved that it worked for every x_{t+1}.

(b) Assume that $\overrightarrow{\mathbf{x_0}}$ is a unit vector ($||\overrightarrow{\mathbf{x_0}}|| = 1$) whose direction is chosen uniformly at random (this basically means that all the possible directions for $\overrightarrow{\mathbf{x_0}}$ are equally likely to be chosen). Let $(\alpha_1, ..., \alpha_n)$ be the coordinates of $\overrightarrow{\mathbf{x_0}}$ in the basis $(\overrightarrow{\mathbf{v_1}}, ..., \overrightarrow{\mathbf{v_n}})$. Explain we have $\alpha_1 \neq 0$ with probability 1. You do not have to do a rigorous proof of that, just give an intuitive argument.

Given a basis $(\overrightarrow{\mathbf{v_1}}, ..., \overrightarrow{\mathbf{v_n}})$, we know that $\overrightarrow{\mathbf{x_0}}$ lies somewhere on an n-dimensional hyper-sphere of radius one. For example, in $\mathbb{R}^3$, $\overrightarrow{\mathbf{x_0}}$ would lie somewhere on the unit sphere. Now, if α_1 were to be zero, this would imply that $\overrightarrow{\mathbf{x_0}}$ in $(\overrightarrow{\mathbf{v_1}}, ..., \overrightarrow{\mathbf{v_n}})$ has no component along $\overrightarrow{\mathbf{v_1}}$. In other words, $\overrightarrow{\mathbf{x_0}}$ would be relegated to an $(n-1)$-dimensional subspace, because the vectors in $(\overrightarrow{\mathbf{v_1}}, ..., \overrightarrow{\mathbf{v_n}})$ are linearly independent. In $\mathbb{R}^3$, $\overrightarrow{\mathbf{x_0}}$ would be bound to a plane, which intersects the hyper-sphere. $\overrightarrow{\mathbf{x_0}}$ would therefore have to lie along the intersection of the sphere and the plane. This region has zero area, and therefore, the probability of lying in this region is zero. In the abstract this reasoning holds, though has no easily interpreted geometric analog. The probability of a vector lying at the intersection of a n-dimensional space and an $(n-1)$-dimensional subspace will always be zero. Therefore $\alpha_1 \neq 0$ with probability 1.

(c) Show that as $t \to \infty$,

$$x_t \to \frac{\alpha_1 \overrightarrow{\mathbf{v_1}}}{||\alpha_1 \overrightarrow{\mathbf{v_1}}||} \qquad ||Ax_t|| \to \lambda_1$$

We know that as $t \to \infty$, we have (vector indicators removed for convenience),

$$\begin{aligned} x_t &= \frac{A^t x_0}{||A^t x_0||} = \frac{A^t(\alpha_1 v_1 + ... + \alpha_n v_n)}{||A^t(\alpha_1 v_1 + ... + \alpha_n v_n)||} = \frac{(\alpha_1 \lambda_1^t v_1 + ... + \alpha_n \lambda_n^t v_n)}{||(\alpha_1 \lambda_1^t v_1 + ... + \alpha_n \lambda_n^t v_n)||} \\ &= \frac{\alpha_1 \lambda_1^t (v_1 + ... + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^t v_n)}{\alpha_1 \lambda_1^t ||(v_1 + ... + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda_n}{\lambda_1}\right)^t v_n)||} \to \frac{\alpha_1 \overrightarrow{\mathbf{v_1}}}{||\alpha_1 \overrightarrow{\mathbf{v_1}}||} \end{aligned} \tag{23}$$

Where the we can make the final simplification since we know $\lambda_1 > |\lambda_i|$. Now, given that we found the value of x_t as $t \to \infty$, we see that at $t \to \infty$,

$$||Ax_t|| \rightarrow ||A\frac{\alpha_1\overrightarrow{\mathbf{v}_1}}{||\alpha_1\overrightarrow{\mathbf{v}_1}||}|| = \frac{\alpha_1||A\overrightarrow{\mathbf{v}_1}||}{\alpha_1||\overrightarrow{\mathbf{v}_1}||} = \frac{||\lambda_1\overrightarrow{\mathbf{v}_1}||}{||\overrightarrow{\mathbf{v}_1}||} = \lambda_1 \tag{24}$$