Assignment 11 – Part 1 – Math 411

(1) (Projecting onto the column space of a matrix.) Let $A \in M_{m,n}(\mathbf{R})$, so that $W = \mathrm{Col}(A) \leq V = \mathbf{R}^m$. We want to write down a formula for the orthogonal projection $P_W : V \to W$ in terms of A.

(a) First off, for a general inner product space V and $S \subseteq V$, let $W = \mathrm{Span}(S)$. Show that $v \in V$ is orthogonal to all $w \in W$ if and only if it is orthogonal to all $w \in S$.

(b) Now, let $V = \mathbf{R}^m$ (with the standard inner product) and $A \in M_{m,n}(\mathbf{R})$. Every vector in $\mathrm{Col}(A)$ is of the form Ax for some $x \in \mathbf{R}^n$, so given $v \in V$, the orthogonal projection onto $\mathrm{Col}(A)$ we seek is $\mathrm{proj}_W v = A\hat{x}$ for some $\hat{x} \in \mathbf{R}^n$ satisfying $v - A\hat{x}$ is orthogonal to all $w \in \mathrm{Col}(A)$. Show that $A\hat{x} = \mathrm{proj}_W v$ if $A^{\mathrm{T}}(v - A\hat{x}) = 0$. (Hint: use part (a) and the fact that $v_1 \cdot v_2 = v_1^{\mathrm{T}} v_2$.)

(c) Suppose the columns of A are linearly independent, and explain why $A^{\mathrm{T}}A$ is invertible.

(d) Suppose the columns of A are linearly independent, and show that $P_W = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ (i.e. for all $v \in V$, we have $\mathrm{proj}_W v = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}v$).

(2) The exercise explains a general application of inner product spaces. Let V be an inner product space and let W be a subspace of V. Given $v \in V$, we seek the closest vector in W to v. Accordingly, we say that $\overline{v} \in W$ is a *best approximation* to v in W if

$$||v - \overline{v}|| \leq ||v - w||, \text{ for all } w \in W.$$

(a) The first step (carried out in parts (a)–(d)) is to show that $\overline{v}$ is a best approximation to v if and only if $v - \overline{v}$ is orthogonal to all vectors in W (in other words, a best approximation is just what in class we called an orthogonal projection of v onto W). To show this, first suppose that $w \in W$ is such that $v - w$ is orthogonal to all vectors in W, and show that $||v - w|| \leq ||v - w'||$ for all $w' \in W$. (Hint: write $v - w'$ as $(v - w) - (w - w')$.)

(b) Given a vector $w' \in W$, show that every vector w'' in W can be written as $w'' = w' - \widetilde{w}$ with $\widetilde{w} \in W$.

(c) Suppose that $||v - w|| \leq ||v - w'||$ for all $w' \in W$. Show that this implies that for all $w'' \in W$, one has $2\mathrm{Re}(\langle v - w, w'' \rangle) + \langle w'', w'' \rangle \geq 0$.

(d) Plugging in $w'' = -\dfrac{\langle v-w, w-w'\rangle}{\langle w-w', w-w'\rangle}(w-w')$ into the inequality of part (c), conclude that if $||v-w|| \leq ||v-w'||$ for all $w' \in W$, then $\langle v-w, w-w'\rangle = 0$ and hence that $v-w$ is orthogonal to all vectors in W.

(e) Show that if a best approximation exists, then it is unique.

(3) Here's an example of applying the idea of a best approximation. The setup is the following. Let $m, n \geq 1$ and let $A \in M_{m,n}(\mathbf{R})$. Let $V = \mathbf{R}^m$ (equipped with the standard inner product) and let $W = \mathrm{Col}(A) \leq V$. For $b \in V$, recall that there is a solution to $Ax = b$ if and only if $b \in \mathrm{Col}(A)$. Sometimes, one has a vector $b \in V$ for which there is no solution to $Ax = b$, but there "should" be (like if b represents an experimental measurement, which should theoretically lie in the column space of A, but doesn't not due to the noise in the experiment). In such a case, we seek $\hat{x} \in \mathbf{R}^n$ such that $||A\hat{x} - b|| \leq ||Ax - b||$ for all $x \in \mathbf{R}^n$. Such a solution is called a *least squares solution* to $Ax = b$.

(a) Let $p = \mathrm{proj}_W b$ and let $\hat{x}$ be a solution of $Ax = p$. Using Question (2) (or otherwise) show that $\hat{x}$ is a least squares solution to $Ax = b$. (Hint: every element of W is of the form Ax for some $x \in \mathbf{R}^n$.)

(b) Show that the only least squares solutions of $Ax = b$ are the solutions to $Ax = p$.

(c) Conclude from Question (1) that the least squares solutions to $Ax = b$ are the solutions to $A^{\mathrm{T}}Ax = A^{\mathrm{T}}b$.

(4) (Least squares linear fit.) Finally, let's give a concrete application. Suppose you have made m measurements of some physical quantity at times $t_1, \cdots, t_m$ and you obtained the values $y_1, \ldots, y_m$, respectively. Suppose that physics tells you there should be a linear relation between y and t, i.e. there's a physical law that says that $y(t)$ satisfies $y(t) = \alpha t + \beta$ for some $\alpha, \beta \in \mathbf{R}$. How do you find α and β given the m data points? Well, one method is the least squares fit. In a perfect world, you would have that $y_i = \alpha t_i + \beta$ for all $i = 1, \ldots, m$, so using two different times would yield α and β. Sadly, the world is far from perfect. Instead, you end up with m equations

$$\begin{aligned} y_1 &= \alpha t_1 + \beta \\ &\vdots \\ y_m &= \alpha t_m + \beta \end{aligned}$$

in the two unknowns α and β, and there's is probably no solution at all! Let

$$b = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \in \mathbf{R}^m \text{ and } A = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{pmatrix} \in M_{m,2}(\mathbf{R}).$$

The *least squares linear fit* of the data points $\{(t_1, y_1), \ldots, (t_m, y_m)\}$ is $y(t) = \alpha t + \beta$, where $\hat{x} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ is the least squares solution to $Ax = b$.

(a) Suppose you have made three measurements $\{(0,0), (1,102), (4,400)\}$. Find the least squares linear fit of this data.