Math 242 fall 2008 notes on problem session for week of 11-3-08
This is a short overview of problems that we covered.

1. Derive the normal equations for the leats squares solution to the system $A\boldsymbol{x} = \boldsymbol{b}$ assuming the geometry (shown in Theorem 5.39 in the text) that the closest point $\boldsymbol{b}$ in a subspace W to $\boldsymbol{v}$ is the orthogonal projection of $\boldsymbol{b}$ onto W. Here W is the subspace spanned by the columns of A and we can write $\boldsymbol{w} = A\boldsymbol{x}^*$ for some $\boldsymbol{x}^*$. So the error vector $\boldsymbol{b} - A\boldsymbol{x}^*$ is orthogonal to W. It is orthogonal if it is orthogonal to every vector in a basis for W. So we want $\boldsymbol{b} - A\boldsymbol{x}^*$ to be orthogonal to every column of A. So we find $\boldsymbol{x}^*$ by solving $A^T(\boldsymbol{b} - A\boldsymbol{x}) = \boldsymbol{0}$. This is $A^T\boldsymbol{b} - A^TA\boldsymbol{x} = \boldsymbol{0}$ which becomes the normal equations $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$. We can let $\boldsymbol{x}^*$ be any solution to these. In particular if $(A^TA)^{-1}$ exists we have $\boldsymbol{x}^* = (A^TA)^{-1}A^T\boldsymbol{b}$.

2. Find the distance from the point $\boldsymbol{x}^T = (x, y)$ in $\mathbb{R}^2$ to the line (through the origin) $ax + by = 0$. We can do this two ways. The first (probably harder) way is to find the error vector $\boldsymbol{e}$ between $\boldsymbol{x}$ and the projection of $\boldsymbol{x}$ onto the line and then find its length. Alternatively we we can project $\boldsymbol{x}$ onto the normal to the line. This gives the same error vector.

   For the first method we let $\boldsymbol{a}^T = (-b, a)$ be a vector in the direction of the line. The projection is $\boldsymbol{p} = \frac{\boldsymbol{a}^T\boldsymbol{x}}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}$. Then using the Pythagorean Theorem we get that $\|\boldsymbol{e}\|^2 = \|\boldsymbol{x}\|^2 - \|\boldsymbol{p}\|^2$. Use $\boldsymbol{x}^T\boldsymbol{x} = (x, y)\begin{pmatrix} x \\ y \end{pmatrix} = (x^2 + y^2)$ and $\boldsymbol{a}^T\boldsymbol{a} = (-b, a)\begin{pmatrix} -b \\ a \end{pmatrix} = (a^2 + b^2)$ and $\boldsymbol{a}^T\boldsymbol{x} = (-b, a)\begin{pmatrix} x \\ y \end{pmatrix} = (ay - bx)$. Then $\|\boldsymbol{e}\|^2 = \boldsymbol{x}^T\boldsymbol{x} - \left(\frac{\boldsymbol{a}^T\boldsymbol{x}}{\boldsymbol{a}^T\boldsymbol{a}}\right)^2 \boldsymbol{a}^T\boldsymbol{a} = \frac{(\boldsymbol{a}^T\boldsymbol{a})(\boldsymbol{x}^T\boldsymbol{x}) - (\boldsymbol{a}^T\boldsymbol{x})^2}{\boldsymbol{a}^T\boldsymbol{a}} = \frac{(a^2+b^2)(x^2+y^2)-(ay-bx)^2}{a^2+b^2} = \frac{(a^2x^2+b^2x^2+a^2y^2+b^2y^2)-(a^2y^2-b^2x^2-2axby)}{a^2+b^2} = \frac{a^2x^2+b^2y^2+2axby}{a^2+b^2} = \frac{(ax+by)^2}{a^2+b^2}$. This is more familiar (for example from calculus) using (x_0, y_0) instead of (x, y) and taking the square root for the distance. Then the distance is $\frac{|ax_0+by_0|}{\sqrt{a^2+b^2}}$.

   For the second method we project onto the direction $\boldsymbol{a}^T = (a, b)$ normal to the line. (This $\boldsymbol{a}$ is different from that in the previous paragraph.) For this $\boldsymbol{a}$ we have $\boldsymbol{a}^T\boldsymbol{a} = (a^2 + b^2)$ and $\boldsymbol{a}^T\boldsymbol{x} = (ax + by)$. The projection is $\frac{\boldsymbol{a}^T\boldsymbol{x}}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}$ and the square of its length is $\left(\frac{\boldsymbol{a}^T\boldsymbol{x}}{\boldsymbol{a}^T\boldsymbol{a}}\right)^2 \boldsymbol{a}^T\boldsymbol{a} = \frac{(\boldsymbol{a}^T\boldsymbol{x})^2}{\boldsymbol{a}^T\boldsymbol{a}} = \frac{(ax+by)^2}{a^2+b^2}$. We get the same result as the previous paragraph.

3. Find the distance from the point $\boldsymbol{x} = (x, y, z)$ in $\mathbb{R}^3$ to the plane (through the origin) $ax + by + cz = 0$. Do this as above. The square of the distance is the length of the projection of $\boldsymbol{x}$ onto the normal to the plane $\boldsymbol{a} = (a, b, c)$. Here $\boldsymbol{a}^T\boldsymbol{a} = (a^2 + b^2 + c^2)$ and $\boldsymbol{a}^T\boldsymbol{x} = (ax + by + cz)$. The projection is The square of its length is $\left(\frac{\boldsymbol{a}^T\boldsymbol{x}}{\boldsymbol{a}^T\boldsymbol{a}}\right)^2 \boldsymbol{a}^T\boldsymbol{a} = \frac{(\boldsymbol{a}^T\boldsymbol{x})^2}{\boldsymbol{a}^T\boldsymbol{a}} = \frac{(ax+by+cz)^2}{a^2+b^2+c^2}$. Taking the square root and using (x_0, y_0, z_0) we get the more familiar form $\frac{|ax_0+by_0+cz_0|}{\sqrt{a^2+b^2+c^2}}$.

4. Find the distance from the point $\boldsymbol{x} = (x, y, z)$ in $\mathbb{R}^3$ to the plane (through the origin) $ax + by + cz = d$. Rewrite this as $ax + by + c(z - \frac{d}{c}) = 0$. Using new coordinates, this is $ax' + by' + cz' = 0$ where the new x and y are the same and the new z is shifted down by c/d. In the new coordinate system the points (x_0, y_0, z_0) becomes $(x_0', y_0', z_0') = (x, y, z + c/d)$. (Since we shifted z down by c/d the distance to the point in the new system is the old distance increased by c/d.) Substituting into the formula above for distance to a plane through the origin we get the distance as $\frac{ax+by+c(z+c/d)}{\sqrt{a^2+b^2+c^2}} = \frac{|ax_0+by_0+cz_0+d|}{\sqrt{a^2+b^2+c^2}}$.

5. If P is projection matrix (that is, $P^2 = P$) show that $(I - P)$ is also a projection matrix. We have $(I-P)^2 = (I-P)(I-P) = I^2 - PI - IP + P^2 = I - P - P + P = I - P$. We note that this is in a sense a generalization of what we did for the distance formulas above. $(I - P)$ projects onto the subspace orthogonal to the original (the normal direction in the case of a line or plane).

6. Let $P = A(A^TA)^{-1}A^T$ be the projection matrix onto the column space of A. Show that $range(P) = Range(A)$.

   $\boldsymbol{b} \in Range(P) \Rightarrow \boldsymbol{b} = P\boldsymbol{z}$ for some $\boldsymbol{z} \Rightarrow \boldsymbol{b} = A(A^TA)^{-1}A^T\boldsymbol{z} \Rightarrow \boldsymbol{b} = A\boldsymbol{w}$ where $\boldsymbol{w} = (A^TA)^{-1}A^T\boldsymbol{z} \Rightarrow \boldsymbol{b} \in Range(A)$. So $Range(P) \subseteq Range(A)$.

   $\boldsymbol{b} \in Range(A) \Rightarrow \boldsymbol{b} = A\boldsymbol{z}$ for some $\boldsymbol{z} \Rightarrow \boldsymbol{b} = AI\boldsymbol{z} = A((A^TA)^{-1}A^TA)\boldsymbol{z} = P(A\boldsymbol{z}) = P\boldsymbol{b} \Rightarrow \boldsymbol{b} \in Range(P)$. So $Range(A) \subseteq Range(P)$.

   Combining we get So $Range(P) = Range(A)$.

   Note that the second set of implications is simply showing that if $\boldsymbol{b}$ is in the range of A then it is the projection of itself onto the range of A.