18.06 (Fall '13) PSet 5 solutions

Exercise 1. In Section 4.2 of the textbook, you learned that if p is the projection of the vector b onto the line a, then p is characterized by the fact that the line from p to b is perpendicular to p. One might guess that this criterion extends to projections onto subspaces of dimension > 1, but this is incorrect: In this question you'll demonstrate, by example, that this approach leads to infinitely many possible "projections". (The right criterion is that the line from p to b is perpendicular to every column of A.)

a) Let A be an $m \times n$ matrix, and let b be a vector in $\mathbb{R}^m$. We'd like to find the projection of b onto the column space of A. If $p = Ax$ is in the column space of A, show that the equation x must satisfy for the line from b to p to be perpendicular to p is

$$x^T A^T b = x^T A^T Ax.$$

b) Now suppose for example A is the $m \times 2$ matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \cdots & \cdots \\ 0 & 0 \end{pmatrix}.$$

Show that in this case, the above equation is just the equation of a circle. Describe clearly the circle.

We'd like to have a unique projection, not a whole circle's worth of them. Thus we must insist that the line from b to p be perpendicular to the entire column space of A.

Solution.

a) We are asked for the equation that guarantees that $(b - Ax)$ and Ax are orthogonal. The orthogonality means that $(Ax)^T(b - Ax) = 0$. Hence, $(Ax)^T b = (Ax)^T Ax$. It follows that $x^T A^T b = x^T A^T Ax$.

b) It is easy to check that $A^T A = I$, so $x^T A^T b = x^T x$. In coordinates: $x_1 b_1 + x_2 b_2 = x_1^2 + x_2^2$. After massaging the equation we get: $(x_1 - b_1/2)^2 + (x_2 - b_2/2)^2 = (b_1^2 + b_2^2)/4$.

Exercise 2. Do Problem 9 from 4.3. For the closest parabola $b = C + Dt + Et^2$ to the same four points, write down the unsolvable equations $Ax = b$ in three unknowns $x = (C, D, E)$. Set up the three normal equations $A^T A\hat{x} = A^T b$ (solution not required). In Figure 4.9a you are now fitting a parabola to 4 points—what is happening in Figure 4.9b?

Solution. The problem refers to four points $t = (0, 1, 3, 4)$ and $b = (0, 8, 8, 20)$ from the previous problems. Plugging in the four values for t into the parabola we get

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}.$$

Thus, the three equations for C, D, and E are:

$$A^TA\begin{bmatrix}C\\D\\E\end{bmatrix} = Ab, \quad \text{or} \quad \begin{bmatrix}4 & 8 & 26\\8 & 26 & 92\\26 & 92 & 338\end{bmatrix}\begin{bmatrix}C\\D\\E\end{bmatrix} = \begin{bmatrix}36\\112\\400\end{bmatrix}.$$

In Figure 4.9b we are building a projection of a vector in 4D onto a 3D plane.

Exercise 3. Do Problem 10 from 4.3. For the closest cubic $b = C + Dt + Et^2 + Ft^3$ to the same four points, write down the four equations $Ax = b$. Solve them by elimination. In Figure 4.9a this cubic now goes exactly through the points. What are p and e?

Solution. The problem refers to four points $t = (0, 1, 3, 4)$ and $b = (0, 8, 8, 20)$ from the previous problems. Plugging in the four values for t into the cubic we get

$$A = \begin{bmatrix}1 & 0 & 0 & 0\\1 & 1 & 1 & 1\\1 & 3 & 9 & 27\\1 & 4 & 16 & 64\end{bmatrix}.$$

Thus, the four equations are:

$$A^TA\begin{bmatrix}C\\D\\E\\F\end{bmatrix} = Ab, \quad \text{or} \quad \begin{bmatrix}4 & 8 & 26 & 92\\8 & 26 & 92 & 338\\26 & 92 & 338 & 1268\\92 & 338 & 1268 & 4826\end{bmatrix}\begin{bmatrix}C\\D\\E\\F\end{bmatrix} = \begin{bmatrix}36\\112\\400\\1504\end{bmatrix}.$$

The system has a unique solution $C = 0$, $D = 47/3$, $E = -28/3$, $F = 5/3$. Matrix A is invertible, the column space is all the space. Hence, $p = b$ and $e = 0$.

Exercise 4. Do Problem 12 from 4.3. This problem projects $b = (b_1, \dots, b_m)$ onto the line through $a = (1, \dots, 1)$. We solve m equations $ax = b$ in 1 unknown (by least squares).

(a) Solve $a^ta\hat{x} = a^tb$ to show that $\hat{x}$ is the *mean* (the average) of the b's.

(b) Find $e = b - a\hat{x}$ and the *variance* $\|e\|^2$ and the *standard deviation* $\|e\|$.

(c) The horizontal line $\hat{b} = 3$ is closest to $b = (1, 2, 6)$. Check that $p = (3, 3, 3)$ is perpendicular to e and find the 3 by 3 projection matrix P.

Solution.

(a) Plugging in the numbers into the formula we get: $m\hat{x} = b_1 + b_2 + \ldots + b_m$, or $\hat{x}$ is the average of the b's.

(b) $e = (b_1 - \hat{x}, b_2 - \hat{x}, \dots, b_m - \hat{x})$. $\|e\|^2 = (b_1 - \hat{x})^2 + (b_2 - \hat{x})^2 + \ldots + (b_m - \hat{x})^2$. $\|e\| = \sqrt{(b_1 - \hat{x})^2 + (b_2 - \hat{x})^2 + \ldots + (b_m - \hat{x})^2}$.

(c) $e = b - p = (-2, -1, 3)$, $ep = (-2) \cdot 3 + (-1) \cdot 3 + 3 \cdot 3 = 0$.

$$A = \begin{bmatrix}1\\1\\1\end{bmatrix} \qquad P = A(A^TA)^{-1}A^T = \begin{bmatrix}1/3 & 1/3 & 1/3\\1/3 & 1/3 & 1/3\\1/3 & 1/3 & 1/3\end{bmatrix}.$$

Exercise 5. Do Problem 13 from 4.3. First assumption behind least squares: $Ax = b-$*(noise e with mean zero)*. Multiply the error vectors $e = b - Ax$ by $(A^TA)^{-1}A^T$ to get $\hat{x} - x$ on the right. The estimation errors $\hat{x} - x$ also average to zero. The estimates $\hat{x}$ is unbiased.

Solution. $(A^TA)^{-1}A^T(b - Ax) = (A^TA)^{-1}A^Tb - (A^TA)^{-1}A^TAx = \hat{x} - x$. When $e = b - Ax$ averages to 0, so does $\hat{x} - x$.

Exercise 6. Do Problem 4 from 4.4. Give an example of each of the following:

(a) A matrix Q that has orthonormal columns but $QQ^T \neq I$.

(b) Two orthogonal vectors that are not linearly independent.

(c) An orthonormal basis for $\mathbb{R}^3$, including the vector $q_1 = (1, 1, 1)/\sqrt{3}$.

Solution.

(a) Such a matrix has to be non-square. Indeed, for a square matrix $Q^TQ = I$. Hence, $Q^T = Q^{-1}$, and $QQ^T = I$. Here is an example:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

(b) Linear dependency of vectors v and w means that there are numbers a and b (both of them can't be zero) such that $av + bw = 0$. From here $0 = (av + bw)^T(av + bw) = a^2\|v\|^2 + b^2\|w\|^2$, because they are orthogonal. Suppose $a \neq 0$, then $v = 0$. That means, one of the vectors must be the zero vector.

(c) For example, $q_1 = (1, 1, 1)/\sqrt{3}$, $q_1 = (1, -1, 0)/\sqrt{2}$, $q_1 = (1, 1, -2)/\sqrt{6}$.

Exercise 7. Do Problem 18 from 4.4. Find orthogonal vectors A, B, C by Gram-Schmidt from a, b, c:

$$a = (1, -1, 0, 0) \quad b = (0, 1, -1, 0) \quad c = (0, 0, 1, -1).$$

Solution. $A = a = (1, -1, 0, 0)$; $B = b - p = (1/2, 1/2, -1, 0)$; $C = c - p_A - p_B = (1/3, 1/3, 1/3, -1)$.

Exercise 8. Do Problem 37 from 4.4. We know that $P = QQ^T$ is the projection onto the column space of $Q(m$ by $n)$. Now add another column a to produce $A = [Q\ a]$. What is the new orthonormal vector q from Gram-Schmidt: start with a, subtract _____, divide by _____.

To rephrase: Q has orthonormal columns. We want to perform Gram-Schmidt on

$$[Q\ a]$$

and we only need to change the final column.

Solution. Start with a, subtract the projection Pa, divide by the length of the result.

Exercise 9. Use Julia or otherwise to compute the coefficients of a best least squares fifth degree approximation to $y = \sin(x)$ on $[0, 2\pi]$.

In Julia you can execute the following code.

```
t=2*pi*(0:.01:1)
A = [t[i]^k for i=1:length(t),  k=0:1:5];
c=float(A)\sin(t)
```

If you would like to see the approximation, you can evaluate the polynomial and plot it:

```
x=(0:.001:1)*2*pi
z=0*x;
for i=length(c):-1:1
  z=z.*x+c[i];
end

using PyPlot
plot(x,z)
plot(x,sin(x))
```

Solution. N/A

Exercise 10. Compare the quintic above to the best solution obtainable from a Taylor series expansion of $\sin x$: $x - x^3/6 + x^5/120$. Also compare with the Taylor series about $x = \pi$: $-(x - \pi) + (x - \pi)^3/6 - (x - \pi)^5/120$.

Solution. The sin function is symmetric on the interval from 0 to 2π with respect to 180° rotation around the point $(\pi, 0)$. The Taylor series are designed to approximate functions locally. So the first expansion would be a good approximation around $x = 0$, but not good overall, as it does not respect the symmetry. The second Taylor series is a good approximation around $x = \pi$. In addition, the series respect the symmetry, so overall it is a much better approximation.