18.06 Problem Set 5
Due Wednesday, 19 March 2008 at 4 pm in 2-106.

Problem 1: Do problem 12 from section 4.3 (pg. 217) in the book. There is a typo in part c. It should read something like this:

    c) Let's analyze what happens when $b = (1, 2, 6)$. In this case $\widehat{x} = 3$ and the projection onto the line is $p = (3, 3, 3)$. Check that p is perpendicular to e. Also find the projection matrix P.

$\boxed{\text{Solution}}$ (4+3+3 points)

    a) We know the vectors a and b, so we can just write this down: $a^T a = a \cdot a = m$, $a^T b = a \cdot b = \sum_i b_i$, so $\widehat{x} = \frac{1}{m} \sum_i b_i$. This is the average of the b_i.

    b) I'll continue to use $\widehat{x}$ to denote the average of the b_i. The error vector is $e = (b_1 - \widehat{x}, b_2 - \widehat{x}, \ldots)$. The variance is then

$$
\begin{aligned}
\|e\|^2 &= \sum_i (b_i - \widehat{x})^2 \\
&= \sum_i b_i^2 - 2\widehat{x} \sum_i b_i + m\widehat{x}^2 \\
&= \sum_i b_i^2 - m\widehat{x}^2 \\
&= \sum_i \left(b_i^2 - \widehat{x}^2 \right)
\end{aligned}
$$

which is the usual expression. The standard deviation is the square root of this quantity.

    c) We have $b = (1, 2, 6)$ and $p = (3, 3, 3)$. The error is $e = (-2, -1, 3)$, and $p \cdot e = 0$ as it should. The matrix P is

$$
P = a(a^T a)^{-1} a^T = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}
$$

Problem 2: Do problem 17 from section 4.3 (pg. 217) in the book.

$\boxed{\text{Solution}}$ (10 points)

The points are $(-1, 7)$, $(1, 7)$, and $(2, 21)$. So, our matrices are

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$b = \begin{bmatrix} 7 \\ 7 \\ 21 \end{bmatrix}$$

We find the coeffiecients (C, D) by setting

$$\begin{bmatrix} C \\ D \end{bmatrix} = \widehat{x} = (A^T A)^{-1} A^T b$$

$$= \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 35 \\ 42 \end{bmatrix}$$

$$= \frac{1}{14} \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 35 \\ 42 \end{bmatrix}$$

$$= \begin{bmatrix} 9 \\ 4 \end{bmatrix}$$

So the best fit line is $b = 9 + 4t$. When you draw it, it will look about right; I'm not going to draw it here.

Problem 3: Find a function of the form $f(t) = C \sin(t) + D \cos(t)$ that approximates the three points $(0, 0)$, $(\pi/2, 2)$, and $(\pi, 1)$. As explained in the book, the method is the same as for fitting a line using least-squares! (See pg. 212 for a quadratic example.) The difference is that the matrix A we use will no longer have columns with entries 1 and t_i but rather $\sin(t_i)$ and $\cos(t_i)$.

$\boxed{\text{Solution}}$ (10 points)

Our matrix A will have columns given by $\sin(t_i)$ and $\cos(t_i)$, but everything else is the same.

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

We find $\widehat{x}$ in the usual way.

$$\begin{aligned}
\widehat{x} &= (A^T A)^{-1} A^T b \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} 2 \\ -\frac{1}{2} \end{bmatrix}
\end{aligned}$$

Our best fit equation is $f(t) = 2\sin(t) - \frac{1}{2}\cos(t)$. Again, if you draw the graph, you'll see that it looks about right.

Problem 4: a) Show that if Q is orthogonal (i.e. Q is square with orthonormal columns) then so is Q^T. Use the criteria $Q^T Q = I$.

b) If Q_1 and Q_2 are orthogonal, show that their product $Q_1 Q_2$ is as well.

c) If Q has orthonormal columns but is not square, does Q^T have the same property? What can you say about QQ^T and $Q^T Q$ in this case? Give an example for a 3 by 2 matrix Q.

$\boxed{\text{Solution}}$ (3+3+4 points)

a) If Q is square, then so is Q^T. So, we just need to show that $(Q^T)^T = (Q^T)^{-1}$. Multiplying: $(Q^T)^T Q^T = QQ^T$. But this is the identity since $Q^T = Q^{-1}$ (as Q is an orthogonal matrix).

b) The same idea here: we must show that $(Q_1 Q_2)^T = (Q_1 Q_2)^{-1}$. Multiplying: $(Q_1 Q_2)^T (Q_1 Q_2) = Q_2^T Q_1^T Q_1 Q_2 = Q_2^T Q_2 = I$. Here we have used that $Q_1^T Q_1 = I$ and $Q_2^T Q_2 = I$.

c) It is not necessarily true that Q^T also has orthonormal columns - this is a very special property of square matrices. For example, suppose

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} \end{bmatrix}$$

Q has orthonormal columns, but its transpose definitely does not (columns 2 and 3 of Q^T are not even orthogonal). In general, any matrix with orthonormal columns

satisfies $Q^T Q = I$, but we can't say anything useful about QQ^T when Q is not square. For example, in this case

$$QQ^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{\sqrt{3}}{4} \\ 0 & \frac{\sqrt{3}}{4} & \frac{3}{4} \end{bmatrix}$$

Problem 5: Do problem 15 from section 4.4 (pg. 229). (Hint for part c: it may be easiest to use the decomposition $A = QR$.)

Solution (10 points)

a) One way to do this is to use Gram-Schmidt on A to find q_1, q_2, and then to find a perpendicular unit vector q_3. Or, we can do it all in one step by adding another column to A - we just need to keep in mind that only q_1 and q_2 will give the Q for A, and not q_3 as well.

I'll use the latter method; I'll set $(1, 0, 0)$ to be the third vector. Running Gram-Schmidt:

$$w_1 = (1, 2, -2)$$

$$\begin{aligned} w_2 &= v_2 - \frac{v_2 \cdot w_1}{w_1 \cdot w_1} w_1 \\ &= (1, -1, 4) - \frac{-9}{9}(1, 2, -2) \\ &= (2, 1, 2) \end{aligned}$$

$$\begin{aligned} w_3 &= v_3 - \frac{v_3 \cdot w_1}{w_1 \cdot w_1} w_1 - \frac{v_3 \cdot w_2}{w_2 \cdot w_2} w_2 \\ &= (1, 0, 0) - \frac{1}{9}(1, 2, -2) - \frac{2}{9}(2, 1, 2) \\ &= \left(\frac{4}{9}, \frac{-4}{9}, \frac{-2}{9} \right) \end{aligned}$$

Finally, we renormalize to find Q:

$$q_1 = \frac{1}{3}(1, 2, -2)$$

$$q_2 = \frac{1}{3}(2, 1, 2)$$

4

$$q_3 = \frac{1}{3}(2, -2, -1)$$

b) By running Gram-Schmidt, we have found a q_3 perpendicular to q_1 and q_2. This means that q_3 is perpendicular to the column space of Q, which is the same as the column space of A. We conclude that q_3 is in the left-nullspace of A.

c) We must solve $Ax = (1, 2, 7)$ by least squares. We find:

$$
\begin{aligned}
\widehat{x} &= (A^T A)^{-1} A^T b \\
&= \begin{bmatrix} 9 & -9 \\ -9 & 18 \end{bmatrix}^{-1} \begin{bmatrix} -9 \\ 27 \end{bmatrix} \\
&= \frac{1}{9} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -9 \\ 27 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ 2 \end{bmatrix}
\end{aligned}
$$

Problem 6: Do problem 18 from section 4.4 (pg. 230). Then write down the $A = QR$ decomposition for the matrix A with columns the given vectors.

Solution (10 points)

We need to find orthogonal vectors by running Gram-Schmidt on

$$
A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}
$$

So, we find

$$w_1 = (1, -1, 0, 0)$$

$$
\begin{aligned}
w_2 &= v_2 - \frac{v_2 \cdot w_1}{w_1 \cdot w_1} w_1 \\
&= (0, 1, -1, 0) - \frac{-1}{2}(1, -1, 0, 0) \\
&= \left(\frac{1}{2}, \frac{1}{2}, -1, 0 \right)
\end{aligned}
$$

$$w_3 = v_3 - \frac{v_3 \cdot w_1}{w_1 \cdot w_1} w_1 - \frac{v_3 \cdot w_2}{w_2 \cdot w_2} w_2$$

$$= (0,0,1,-1) - \frac{0}{2}(1,-1,0,0) - \frac{-1}{3/2}\left(\frac{1}{2},\frac{1}{2},-1,0\right)$$

$$= \left(\frac{1}{3},\frac{1}{3},\frac{1}{3},-1\right)$$

Finally, we renormalize to find Q:

$$q_1 = \frac{1}{\sqrt{2}}(1,-1,0,0)$$

$$q_2 = \frac{\sqrt{2}}{\sqrt{3}}(\frac{1}{2},\frac{1}{2},-1,0)$$

$$q_3 = \frac{\sqrt{3}}{2}(\frac{1}{3},\frac{1}{3},\frac{1}{3},-1)$$

We also need to find R. From the equation $A = QR$, we find $R = Q^T A$. Multiplying, we get

$$R = \begin{bmatrix} \sqrt{2} & \frac{-1}{\sqrt{2}} & 0 \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} & \frac{-\sqrt{2}}{\sqrt{3}} \\ 0 & 0 & \frac{2}{\sqrt{3}} \end{bmatrix}$$

Problem 7: Do problem 3 in section 5.1 (pg. 240).

Solution (10 points)

a) False. A counterexample is given by

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

Then $\det(A) = 1$ but $\det(A + I) = 0$.

b) True. This is just the product rule for determinants applied twice.

c) False. If A has dimension n, the correct answer would be $\det(4A) = 4^n \det(A)$. So the 2 by 2 identity is a counterexample.

Be careful here - it *is* true that the determinant is linear in each row separately. So if we multiplied just one row by 4, we would multiply the determinant by 4 as

6

well. But if we multiply *every* row by 4, we must multiply the determinant by one factor of 4 for each row.

d) False. An example is given by

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

so that

$$AB - BA = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

which has determinant -1.

Problem 8: Do problem 8 in section 5.1 (pg. 241).

Solution (7+3 points)
   a) We have

$$\det(Q)\det(Q) = \det(Q^T)\det(Q) = \det(Q^TQ) = \det(I) = 1$$

Since $\det(Q)^2 = 1$, we must have $\det(Q) = \pm 1$.

b) This problem is a bit tricky. The proper way to approach it is using eigenvalues and eigenvectors, but of course we haven't talked about those yet.
   We need to show that we can't have $|\det(Q)| > 1$. (The problem doesn't ask about the situation when $|\det(Q)| < 1$, so we are not going to worry about it now.) If the determinant were larger than 1, then the product rule tells us that $|\det(Q^n)|$ can be made arbitrarily large if we choose n large enough. However, we know that the length of a vector is not changed by Q, that is $\|Qx\| = \|x\|$. So, if we can somehow compare the size of $|\det(Q)|$ to the length of vectors $\|Qx\|$, we can come up with a contradiction.
   This relationship is most precisely described by eigenvalues, but here's another way. Suppose that A is any invertible matrix. We know that $|\det(A)|$ is the volume of the paralleliped formed by the columns of A (the volume will be non-zero since the system is non-degenerate). Now, suppose that we repeatedly act on these column vectors by Q. First, recall that Q doesn't change the length of any vectors. So, the largest volume that we can end up with (no matter how many times we multiply

7

by Q) is the product of the lengths of the columns of A. The volume can't keep increasing - it is bounded by some number.

However, each time we act on our vectors by Q we increase the volume of the paralleliped by a factor of $|\det(Q)|$. If $|\det(Q)| > 1$, then the volume of our paralleliped would become larger and larger as we act on it by Q more and more. This is a contradiction to the boundedness we just found above.

Problem 9: Find the determinants of the following matrices (Use determinant = product of pivots and/or row swaps.)

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 5 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 4 & 0 & 0 & 6 \\ 0 & 3 & 0 & 0 \end{bmatrix}$$

$\boxed{\text{Solution}}$ (10 points)

We find the determinant of A by reducing and then calculating the determinant of U:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

The determinant of U is the product of the diagonal entries, so $\det(A) = \det(U) = 1$.

We find the determinant of B by doing row swaps. The goal will be to obtain a lower triangular matrix; we can then find the determinant simply by multiplying the pivots. Each row swap introduces a factor of -1 which we need to keep track

of.

$$
\det \begin{bmatrix} 0 & 5 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 4 & 0 & 0 & 6 \\ 0 & 3 & 0 & 0 \end{bmatrix} = (-1)\det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 5 & 2 & 0 \\ 4 & 0 & 0 & 6 \\ 0 & 3 & 0 & 0 \end{bmatrix}
$$

$$
= (-1)^2 \det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 4 & 0 & 0 & 6 \\ 0 & 5 & 2 & 0 \end{bmatrix}
$$

$$
= (-1)^3 \det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 5 & 2 & 0 \\ 4 & 0 & 0 & 6 \end{bmatrix}
$$

So, the determinant of B is -36.

Problem 10: Here's another Matlab question. You can download the code from the "Problem Sets" section of the webpage, or just refer to the commands below. This problem concerns the condition number c which measures the sensitivity to errors of an equation $Ax = b$. (The condition number is $c = \|A\|\,\|A^{-1}\|$ - see pg. 462 for details.) Generally you lose log c decimal places to roundoff in solving Ax=b.

This problem came to me from a GPS expert who noticed that in this least-squares problem the condition number *increases* with more observations. We fit a function of the form $C + Dt + E\sin(2\pi t)$ to a set of data points that are evenly spaced over an interval from 0 to tmax. The Matlab code will output the condition number for different tmax and observation numbers.

a) What is the 3 by 3 matrix A with 3 observations that has such a big condition number on the first line of the output?

b) What condition numbers do you get if tmax is reduced to .02? Use $3, 4, 5, 10, 100$ observations as in the code.

Matlab code:

```
function fitlinesine
obs=[3,4,5,10,100]; % numbers of observations to use
tmax=[1 1/2 1/4 1/8 1/16]; % maximum values of t to use
for i=1:length(tmax) % for each tmax
```

```
    tm=tmax(i); % current tmax
    disp(sprintf('For the interval [0,%f]:',tm));
    for j=1:length(obs) % for each number of observations
      m=obs(j); % current number of observations (will be number of rows)
      t=(0:m-1)'*tm/(m-1); % m linearly spaced t's from 0 to tmax
      A=[ones(m,1) t sin(2*pi*t)]; % construct matrix A
      c=cond(A); % indicates how ill conditioned system is
      disp(sprintf('%5d observations : cond=%f',m,c));
    end
    disp(' ');
end
```

Solution (10 points)

The output is copied at the end; I edited the code slightly so that it would include the case when tmax $= 0.02$ (just add this number to the vector tmax).

a) The rows of A are given by $(1, t, \sin(2\pi t))$ for the points t at which there is an observation. Here we have split the interval from 0 to 1 into three pieces, so we obtain three observations and three rows, one at $t = 0$, $t = 1/2$, and $t = 1$. Thus the matrix is

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1/2 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

The first column corresponds to the constant term C, the next is the coefficient of the linear term Dt, and the last is the coefficient of $E\sin(2\pi t)$. Note that this A is not invertible, which leads to the high condition number.

b) The condition numbers make up the last section of the output. These numbers are very high, which indicate that it is difficult to weight the straight line coefficient D and the sin coefficient E. This occurs since we are on a very small interval close to 0, where the sin function looks almost exactly the same as a straight line (look at the Taylor series).

```
For the interval [0,1.000000]:
    3 observations : cond=9826207312743670.000000
    4 observations : cond=3.686771
    5 observations : cond=4.151688
   10 observations : cond=5.393387
  100 observations : cond=7.058335
```

10

```
For the interval [0,0.500000]:
     3 observations : cond=5.545108
     4 observations : cond=6.281994
     5 observations : cond=6.742934
    10 observations : cond=7.714562
   100 observations : cond=8.660958

For the interval [0,0.250000]:
     3 observations : cond=51.074911
     4 observations : cond=55.064347
     5 observations : cond=58.910363
    10 observations : cond=69.360809
   100 observations : cond=82.322159

For the interval [0,0.125000]:
     3 observations : cond=448.483426
     4 observations : cond=472.857269
     5 observations : cond=501.776602
    10 observations : cond=585.291552
   100 observations : cond=692.212187

For the interval [0,0.062500]:
     3 observations : cond=3578.889606
     4 observations : cond=3758.900278
     5 observations : cond=3982.988668
    10 observations : cond=4637.934580
   100 observations : cond=5481.419212

For the interval [0,0.020000]:
     3 observations : cond=108867.496487
     4 observations : cond=114222.995934
     5 observations : cond=120983.822947
    10 observations : cond=140810.452216
   100 observations : cond=166386.674376
```