18.06 Problem Set 5 - Solutions

Due Wednesday, 17 October 2007 at 4 pm in 2-106.

Problem 1: (10) Do problem 22 from section 4.1 (P 193) in your book.

Solution The equation $x_1 + x_2 + x_3 + x_4 = 0$ can be rewritten in the matrix form

$$\begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = 0.$$

Thus $\mathbf{P}$ is the nullspace of the 1 by 4 matrix

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}.$$

This implies that $\mathbf{P}^\perp$ is the row space of A. Obviously a basis of $\mathbf{P}^\perp$ is given by the vector

$$\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Problem 2: (15=6+3+6) (1) Derive the *Fredholm Alternative:* If the system $A\mathbf{x} = \mathbf{b}$ has no solution, then argue there is a vector $\mathbf{y}$ satisfying

$$A^T\mathbf{y} = 0 \text{ with } \mathbf{y}^T\mathbf{b} = 1.$$

(Hint: $\mathbf{b}$ is not in the column space $C(A)$, thus $\mathbf{b}$ is not orthogonal to $N(A^T)$.)

Solution Suppose the system $A\mathbf{x} = \mathbf{b}$ has no solution, in other words, the vector $\mathbf{b}$ does not lie in the column space $C(A)$. Then $\mathbf{b}$ is not orthogonal to the nullspace $N(A^T)$. Let $\mathbf{p}$ be the orthogonal projection of $\mathbf{b}$ onto $N(A^T)$, then $\mathbf{p} \neq 0$. We have

$$\mathbf{p}^T\mathbf{b} = \mathbf{p}^T\mathbf{p} \neq 0.$$

Let $\mathbf{y} = \frac{1}{\mathbf{p}^T\mathbf{p}}\mathbf{p}$, we see that

$$A^T\mathbf{y} = \frac{1}{\mathbf{p}^T\mathbf{p}}A^T\mathbf{p} = 0$$

but

$$\mathbf{y}^T\mathbf{b} = \frac{1}{\mathbf{p}^T\mathbf{p}}\mathbf{p}^T\mathbf{b} = 1.$$

(2) Check that the following system $A\mathbf{x} = \mathbf{b}$ has no solution:

$$\begin{aligned} x + 2y + 2z &= 2 \\ 2x + 2y + 3z &= 1 \\ 3x + 2y + 4z &= 2 \end{aligned}$$

Solution We do Gauss elimination:

$$\begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 2 & 3 & 1 \\ 3 & 2 & 4 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 2 & 2 \\ 0 & -2 & -1 & -3 \\ 0 & -4 & -2 & -4 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 2 & 2 \\ 0 & -2 & -1 & -3 \\ 0 & 0 & 0 & 2 \end{pmatrix},$$

which certainly has no solution.

(3) Find a vector $\mathbf{y}$ for above system such that $A^T\mathbf{y} = 0$ and $\mathbf{y}^T\mathbf{b} = 1$.

Solution From solution to part (1) one need to find the projection of the vector $\mathbf{b}$ onto the $N(A^T)$. We compute $N(A^T)$:

$$A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 2 & 3 \\ 3 & 2 & 4 \end{pmatrix}$$

$$\Rightarrow A^T = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ 2 & 3 & 4 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 3 \\ 0 & -2 & -4 \\ 0 & -1 & -2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 3 \\ 0 & -2 & -4 \\ 0 & 0 & 0 \end{pmatrix}.$$

So the nullspace $N(A^T)$ is spanned by one vector $\mathbf{a} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$.

The projection of $\mathbf{b}$ on the this line is

$$\mathbf{p} = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}\mathbf{a} = \frac{2 - 2 + 2}{1 + 4 + 1}\mathbf{a} = \frac{2}{6}\mathbf{a} = \begin{pmatrix} 1/3 \\ -2/3 \\ 1/3 \end{pmatrix}$$

So the vector $\mathbf{y}$ we need is

$$\mathbf{y} = \frac{1}{\mathbf{p}^T\mathbf{p}}\mathbf{p} = \frac{1}{1/9 + 4/9 + 1/9}\mathbf{p} = \begin{pmatrix} 1/2 \\ -1 \\ 1/2 \end{pmatrix}$$

Problem 3: (10=2+2+2+2+2) Justify the following (true) statements:

(1) If $AB = 0$, then the column space of B is in the nullspace of A.

Solution If not, i.e., there is a vector $\mathbf{y} = B\mathbf{x}$ lies in the column space of B, but not in the nullspace of A. Then

$$(AB)\mathbf{x} = A(B\mathbf{x}) \neq 0,$$

contradicts with $AB = 0$.

(2) If A is symmetric matrix, then its column space is perpendicular to its nullspace.

Solution Since A is symmetric, $A = A^T$. So its columm space coincides with its row space: $C(A) = C(A^T)$. This implies that its column space is perpendicular to its nullspace.

(3) If a subspace S is contained in a subspace V, then $S^\perp$ contains $V^\perp$.

Solution Suppose $\mathbf{v} \in V^\perp$, i.e., $\mathbf{v}$ is perpendicular to any vector in V. In particular, $\mathbf{v}$ is perpendicular to any vector in S, since $S \subset V$. This shows that $\mathbf{v} \in S^\perp$. So $S^\perp \supset V^\perp$.

(4) For any subspace V, $(V^\perp)^\perp = V$.

Solution By definition, $V^\perp$ is the set of vectors that are perpendicular to all vectors in V. So any vector in V is perpendicular to all vectors in $V^\perp$. This implies $V \subset (V^\perp)^\perp$. On the other hand, suppose the dimension of V is r, then the dimension of $V^\perp$ is $n-r$, and the dimension of $(V^\perp)^\perp$ is again r. So a basis of V is also a basis of $(V^\perp)^\perp$. This implies $(V^\perp)^\perp = V$.

(Another way: any subspace V is defined by some linear equations, in other words, $V = N(A)$ is the nullspace for some matrix A. Thus $V^\perp = C(A^T)$ by the fundamental theorm of linear algebra. Use this theorem again we get $(V^\perp)^\perp = N((A^T)^T) = N(A) = V$.)

(The proofs above only work for finite dimensional spaces. However, the statement is true for any closed subspaces in infinitely dimensional vector spaces, and the proof is much harder.)

(5) If P is a projection matrix, so is $I - P$.

Solution Suppose P is the projection matrix onto a subspace V. Then $I - P$ is the projection matrix that projects onto $V^\perp$. In fact, for any vector $\mathbf{v}$,

$$\mathbf{v} - (I - P)\mathbf{v} = \mathbf{v} - \mathbf{v} + P\mathbf{v} = P\mathbf{v},$$

and obviously $P\mathbf{v} \in V$ is perpendicular to $V^\perp$.

Problem 4: (10=5+5) (1) Do problem 5 from section 4.2 (P 203) in your book.

Solution We compute

$$P_1 = \frac{\mathbf{a}_1\mathbf{a}_1^T}{\mathbf{a}_1^T\mathbf{a}_1} = \frac{1}{1+4+4}\begin{pmatrix} 1 & -2 & -2 \\ -2 & 4 & 4 \\ -2 & 4 & 4 \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 1 & -2 & -2 \\ -2 & 4 & 4 \\ -2 & 4 & 4 \end{pmatrix},$$

$$P_2 = \frac{\mathbf{a}_2\mathbf{a}_2^T}{\mathbf{a}_2^T\mathbf{a}_2} = \frac{1}{4+4+1}\begin{pmatrix} 4 & 4 & -2 \\ 4 & 4 & -2 \\ -2 & -2 & 1 \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 4 & 4 & -2 \\ 4 & 4 & -2 \\ -2 & -2 & 1 \end{pmatrix}.$$

Their product is

$$P_1P_2 = \frac{1}{9}\begin{pmatrix} 1 & -2 & -2 \\ -2 & 4 & 4 \\ -2 & 4 & 4 \end{pmatrix}\frac{1}{9}\begin{pmatrix} 4 & 4 & -2 \\ 4 & 4 & -2 \\ -2 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This product is identically zero, since $\mathbf{a}_1$ and $\mathbf{a}_2$ are perpendicular, and thus if we first project a vector onto $\mathbf{a}_1$, then project the projection onto $\mathbf{a}_2$, we will get the zero vector.

(2) Do problem 7 from section 4.2 (P 203) in your book.

Solution The matrix P_3 is

$$P_3 = \frac{\mathbf{a}_3\mathbf{a}_3^T}{\mathbf{a}_3^T\mathbf{a}_3} = \frac{1}{4+1+4}\begin{pmatrix} 4 & -2 & 4 \\ -2 & 1 & -2 \\ 4 & -2 & 4 \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 4 & -2 & 4 \\ -2 & 1 & -2 \\ 4 & -2 & 4 \end{pmatrix}.$$

Obviously that

$$P_1 + P_2 + P_3 = \frac{1}{9}\begin{pmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{pmatrix} = I.$$

Finally we verify that $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ are orthogonal:

$$\mathbf{a}_1^T\mathbf{a}_2 = -2+4-2 = 0;$$
$$\mathbf{a}_1^T\mathbf{a}_3 = -2-2+4 = 0;$$
$$\mathbf{a}_2^T\mathbf{a}_3 = 4-2-2 = 0.$$

Problem 5: (15=5+5+5) (1) Find the projection matrix P_C onto the column space of

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 4 & 8 & 4 \end{pmatrix}.$$

Solution By observation it is easy to see that the column space of A is the one dimensional subspace containing the vector $\mathbf{a} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$. Thus the projection matrix is

$$P_C = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}} = \frac{1}{17}\begin{pmatrix} 1 & 4 \\ 4 & 16 \end{pmatrix}.$$

(2) Find the projection matrix P_R onto the row space of the above matrix.

Solution By observation the row space of the matrix A is the one dimensional subspace containing the vector $\mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$. Thus the projection matrix is

$$P_R = \frac{\mathbf{b}\mathbf{b}^T}{\mathbf{b}^T\mathbf{b}} = \frac{1}{6}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

(3) What is $P_C A P_R$? Explain your result.

Solution We calculate

$$\begin{aligned} P_C A P_R &= \frac{1}{17}\begin{pmatrix} 1 & 4 \\ 4 & 16 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 1 \\ 4 & 8 & 4 \end{pmatrix} \cdot \frac{1}{6}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \\ &= \frac{1}{6}\begin{pmatrix} 1 & 2 & 1 \\ 4 & 8 & 4 \end{pmatrix}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & 1 \\ 4 & 8 & 4 \end{pmatrix} \\ &= A. \end{aligned}$$

For any vector $\mathbf{v}$, we see $\mathbf{v} - P_R\mathbf{v}$ is always perpendicular to the row space of A, thus $\mathbf{v} - P_R\mathbf{v} \in N(A)$. So $A(\mathbf{v} - P_R\mathbf{v}) = 0$, i.e. $A\mathbf{v} = AP_R\mathbf{v}$. This implies $A = AP_R$. Similarly, $A\mathbf{v} \in C(A)$ implies $P_C A\mathbf{v} = A\mathbf{v}$, i.e., $A = P_C A$. So we always have $P_C A P_R = P_C(AP_R) = P_C A = A$.

Problem 6: (10=3+4+3) Do problem 12 from section 4.3 (P 217) in your book.

Solution (a) Since

$$\mathbf{a}^T\mathbf{a} = 1 + 1 + \cdots + 1 = m,$$
$$\mathbf{a}^T\mathbf{b} = b_1 + b_2 + \cdots + b_m$$

We see that the equation $\mathbf{a}^T\mathbf{a}\hat{x} = \mathbf{a}^T\mathbf{b}$ is equivalent to the equation

$$m\hat{x} = b_1 + b_2 + \cdots + b_m,$$

The solution is given by

$$\hat{x} = \frac{b_1 + b_2 + \cdots + b_m}{m},$$

the mean of the $\mathbf{b}$'s.

(b) We calculate:

$$\mathbf{e} = \mathbf{b} - \hat{x}\mathbf{a} = (b_1 - \hat{x}, b_2 - \hat{x}, \cdots, b_m - \hat{x}),$$

where $\hat{x}$ is the mean above. So the variance is

$$\begin{aligned}\|e\|^2 &= (b_1 - \hat{x})^2 + (b_2 - \hat{x})^2 + \cdots + (b_m - \hat{x})^2 \\ &= b_1^2 + b_2^2 + \cdots + b_m^2 - 2(b_1 + b_2 + \cdots + b_m)\hat{x} + m\hat{x}^2 \\ &= b_1^2 + b_2^2 + \cdots + b_m^2 - 2m\hat{x}^2 + m\hat{x}^2 \\ &= b_1^2 + b_2^2 + \cdots + b_m^2 - m\hat{x}^2.\end{aligned}$$

The standard deviation is

$$\|e\| = \sqrt{b_1^2 + b_2^2 + \cdots + b_m^2 - m\hat{x}^2}.$$

(c) Now

$$m = 3, \mathbf{b} = (1, 2, 6).$$

So

$$\hat{x} = \frac{1 + 2 + 6}{3} = 3,$$

and

$$\mathbf{e} = (1, 2, 6) - (3, 3, 3) = (-2, -1, 3).$$

Obviously $\mathbf{p} = (3, 3, 3)$ is perpendicular to $\mathbf{e}$:

$$\mathbf{p}^T\mathbf{e} = -6 - 3 + 9 = 0.$$

Problem 7: (10=5+5) In this problem you will derive *weighted least-squares* fits. In particular, suppose that you have m data points (t_i, b_i), that you want to fit to a line $b = C + Dt$. Ordinary least squares would choose C and D to minimize the sum-of-squares error $\sum_i (C + Dt_i - b_i)^2$, as derived in class. However, not all data points are always created equal: often, real data points come with a margin of error $\sigma_i > 0$ in b_i. When choosing C and D, we want to weight the data points *less* if they have *more* error. In particular, we want to choose C and D to minimize the error ϵ given by:

$$\epsilon = \sum_{i=1}^{m} \left(\frac{C + Dt_i - b_i}{\sigma_i} \right)^2 .$$

(a) Write ϵ in matrix form, just as for ordinary least squares in class (i.e. with a matrix A of 1s and t_i values and a vector $\mathbf{b}$ of b_i values), but using the additional diagonal "weighting" matrix W with $W_{ii} = 1/\sigma_i$ and $W_{ij} = 0$ for $i \neq j$.

Solution In matrix form

$$\epsilon = \|WA\mathbf{x} - W\mathbf{b}\|^2,$$

where

$$A = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}, \quad W = \begin{pmatrix} 1/\sigma_1 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_m. \end{pmatrix}$$

(b) Derive a linear equation whose solution is the 2-component vector $\mathbf{x}$ ($x_1 = C$, $x_2 = D$) minimizing ϵ.

Solution Now we are minimizing

$$\|WA\mathbf{x} - W\mathbf{b}\|^2.$$

This is just the ordinary least square problem with A replaced by WA, and $\mathbf{b}$ replaced by $W\mathbf{b}$. So the linear equation whose solution minimizing ϵ is

$$(WA)^T(WA)\hat{x} = (WA)^T W\mathbf{b},$$

i.e.

$$A^T W^2 A\hat{x} = A^T W^2 \mathbf{b}.$$

More explicitly,

$$\begin{pmatrix} \sum 1/\sigma_i^2 & \sum t_i/\sigma_i^2 \\ \sum t_i/\sigma_i^2 & \sum t_i^2/\sigma_i^2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} \sum b_i/\sigma_i^2 \\ \sum t_i b_i/\sigma_i^2 \end{pmatrix}.$$

Problem 8: (20=4+4+2+5+5) For this problem, you will generate some random data points from $b = C + Dt +$ noise for $C = 1$ and $D = 0.5$, and then try to use least-square fitting to recover C and D.

(a) First, generate m random data points for $m = 20$ and $t \in (0, 10)$:

```
m = 20
t = rand(m,1) * 10
b = 1 + 0.5*t + (rand(m,1)-0.5)
```

The last line generates the data points from $C + Dt$ plus random numbers in $(-0.5, 0.5)$. Plot them with:

```
plot(t, b, 'o')
```

Solution The codes

```
>> m=20;t=rand(m,1)*10,b=1+0.5*t+(rand(m,1)-0.5),plot(t,b,'o')

t =

    4.3874
    3.8156
    7.6552
    7.9520
    1.8687
    4.8976
    4.4559
    6.4631
    7.0936
    7.5469
    2.7603
    6.7970
    6.5510
    1.6261
    1.1900
    4.9836
    9.5974
    3.4039
```

```
    5.8527
    2.2381


b =

    3.4450
    2.6629
    4.8335
    5.1751
    2.3253
    3.9081
    3.2751
    3.8702
    4.1961
    4.5309
    2.7208
    4.1528
    4.5898
    1.5566
    2.0243
    3.3418
    5.4953
    2.4530
    4.0424
    2.0923
```

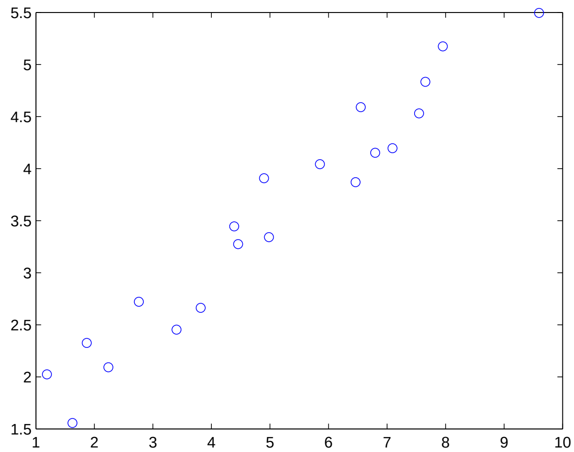


Figure 1: t-b

(b) Now, do the least-square fit, as in class, by constructing the matrix A:

```
A = [ ones(m, 1), t ]
```

and then solving $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$ for $\hat{\mathbf{x}} = (C; D)$:

```
x = (A' * A) \ (A' * b)
```

(Refer to the 18.06 Matlab cheat-sheet if some of these commands confuse you.) Plot the least-square fit, along with the "real" line $1 + t/2$:

```
t0 = [0; 10]
plot(t, b, 'bo', t0, x(1) + t0*x(2), 'r-', t0, 1 + t0/2, 'k--')
```

(The data points should be blue circles, the least-square fit a red line, and the "real" line a black dashed line.)

Solution Codes

```
>> A=[ones(m,1),t];x=(A'*A)\(A'*b),t0=[0;10];
plot(t,b,'bo',t0,x(1)+t0*x(2),'r-',t0,1+t0/2,'k--')

x =

    1.2264
    0.4565
```

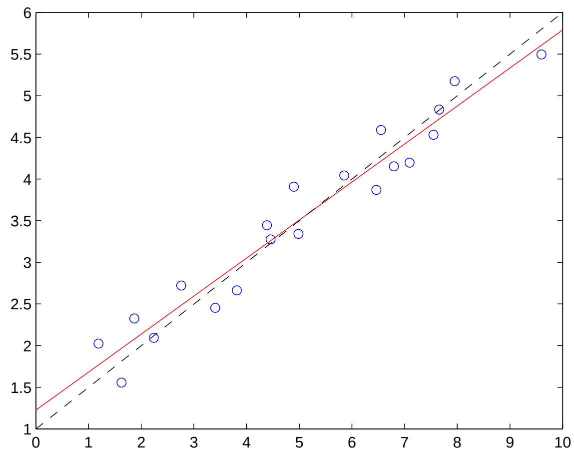


Figure 2: least-square

(c) Verify that you get the same **x** by either of the two commands:

```
x = A \ b
x = pinv(A) * b
```

Solution Codes

```
>> x=A \ b

x =

    1.2264
    0.4565

>> x=pinv(A)*b

x =

    1.2264
    0.4565
```

(d) Repeat the least-square fit process above (you can skip the plots) for increasing numbers of data points: $m = 40, 80, 160, 320, 640, 1280$ (and more, if you want). For each one, compute the squared error E in the least-square C and D compared to their "real" values in the formula that the data is generated from:

```
E = (x(1) - 1)^2 + (x(2) - 0.5)^2
```

Plot this squared error versus m on a log-log scale using the command `loglog` in Matlab (which works just like `plot` but with logarithmic axes). Overall, you should find that the error decreases with m: with more data points, the noise in the data averages out and the fit gets closer and closer to the underlying formula $b = 1 + t/2$. Note that if you want to create an array of E values, you can assign the elements one by one via `E(1) = ...; E(2) = ...;` and so on. (Or you can write a loop, for VI-3 hackers.)

Solution Codes

```
>> m=40;t=rand(m,1)*10;b=1+0.5*t+(rand(m,1)-0.5);A=[ones(m,1),t];
x=(A'*A)\(A'*b);E(1)=(x(1)-1)^2+(x(2)-0.5)^2

E =

    0.0073

>> m=80;t=rand(m,1)*10;b=1+0.5*t+(rand(m,1)-0.5);A=[ones(m,1),t];
x=(A'*A)\(A'*b);E(2)=(x(1)-1)^2+(x(2)-0.5)^2

E =

    0.0073    0.0019

>> m=160;t=rand(m,1)*10;b=1+0.5*t+(rand(m,1)-0.5);A=[ones(m,1),t];
x=(A'*A)\(A'*b);E(3)=(x(1)-1)^2+(x(2)-0.5)^2

E =

    0.0073    0.0019    0.0018

>> m=320;t=rand(m,1)*10;b=1+0.5*t+(rand(m,1)-0.5);A=[ones(m,1),t];
x=(A'*A)\(A'*b);E(4)=(x(1)-1)^2+(x(2)-0.5)^2
```

```
E =

    0.0073    0.0019    0.0018    0.0008

>> m=640;t=rand(m,1)*10;b=1+0.5*t+(rand(m,1)-0.5);A=[ones(m,1),t];
x=(A'*A)\(A'*b);E(5)=(x(1)-1)^2+(x(2)-0.5)^2

E =

    0.0073    0.0019    0.0018    0.0008    0.0004

>> m=1280;t=rand(m,1)*10;b=1+0.5*t+(rand(m,1)-0.5);A=[ones(m,1),t];
x=(A'*A)\(A'*b);E(6)=(x(1)-1)^2+(x(2)-0.5)^2

E =

    0.0073    0.0019    0.0018    0.0008    0.0004    0.0001

>> m(1)=40;m(2)=80;m(3)=160;m(4)=320;m(5)=640;m(6)=1280;loglog(m,E,'bo')
```

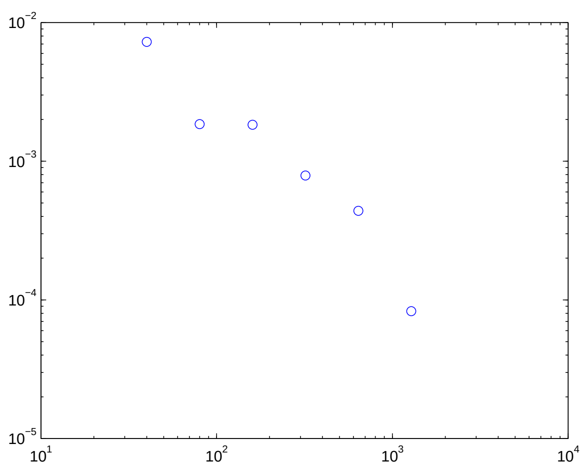


Figure 3: m-E

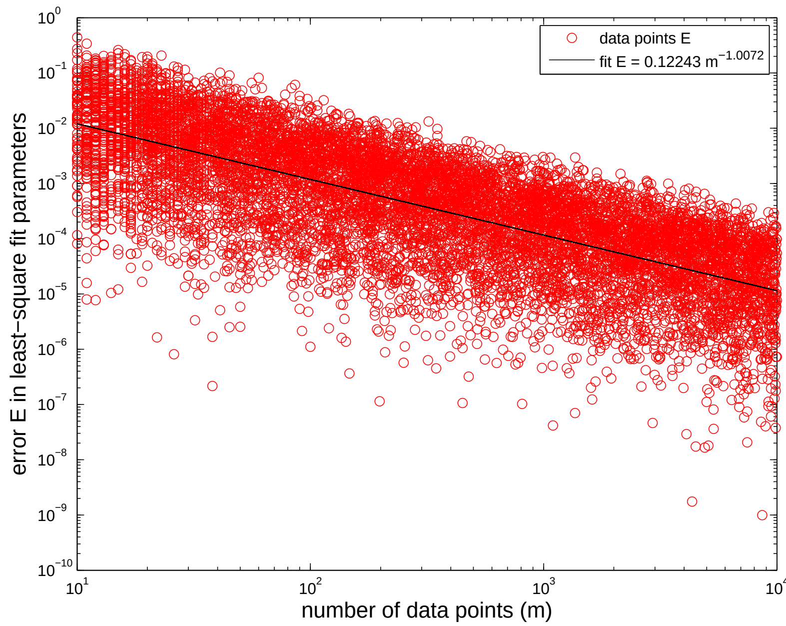
(e) Overall, E should depend on m as some power law: $E = \alpha * m^\beta$ for some constants α and β (plus random noise, of course). Find α and β by a least-square fit of $\log E$ versus $\log m$ (since $\log E = \log \alpha + \beta \log m$ is a straight line). (Show your code!)

Solution Codes

```
>> lm(1)=log(m(1));lm(2)=log(m(2));lm(3)=log(m(3));lm(4)=log(m(4));
lm(5)=log(m(5));lm(6)=log(m(6));

>> le(1)=log(E(1));le(2)=log(E(2));le(3)=log(E(3));le(4)=log(E(4));
le(5)=log(E(5));le(6)=log(E(6));

>> B=[ones(6,1),lm'];y=(B'*B)\(B'*le')

y =

   -0.8137
   -1.1346
```

Thus $\alpha = e^{-0.8137} = 0.4432, \beta = -1.1346$.

(More accurate solution should go to about α=0.12, β=-1. Prof. Johnson tried it for 10000 random m values log-distributed from 10 to 10000 — see the graph below. The actual student answers will vary quite a bit because of random variations, of course (for the suggested data set of only 6 data points, the standard deviation of beta seems to be about 0.7).)



data points E
fit E = 0.12243 m^−1.0072
error E in least−square fit parameters
number of data points (m)
10^0
10^−1
10^−2
10^−3
10^−4
10^−5
10^−6
10^−7
10^−8
10^−9
10^−10
10^1
10^2
10^3
10^4

Figure 4: m-E

For problem 3(iv), we would ideally like to prove $(V^\perp)^\perp = V$ for "any" subspace V without assuming a finite-dimensional vector space. We need to show both $V \subset (V^\perp)^\perp$ and $(V^\perp)^\perp \subset V$:

- If $v \in V$, then v is perpendicular to everything in $V^\perp$, by definition, so $v \in (V^\perp)^\perp$.
- If $y \in (V^\perp)^\perp$, let v be the closest point[1] in V to y, i.e. v is the point in V that minimizes $\|y - v\|^2$—we now must show that $y = v$. In class, we showed $y - v \in V^\perp$ for finite-dimensional spaces, using calculus; if we can show the same thing in general we are done: $y \in (V^\perp)^\perp$ implies that $y = (y - v) + v$ is perpendicular to everything in $V^\perp$, which implies that $y - v$ is perpendicular to everything in $V^\perp$ (since v is perpendicular to $V^\perp$), which implies that $y - v$ is 0 (the only element of $V^\perp$ that is also perpendicular to $V^\perp$), and hence $y = v$.
- To show $y - v \in V^\perp$, consider any point $v' \in V$ and any real number λ (assuming our vector space is over the reals). V is a subspace, so $v + \lambda v' \in V$, and v is the closest point in V to y, so $\|y - v\|^2 \leq \|y - (v + \lambda v')\|^2 = \|y - v\|^2 + \lambda^2 \|v'\|^2 - 2\lambda v' \cdot (y - v)$. Choose the sign of λ so that $\lambda v' \cdot (y - v) = |\lambda v' \cdot (y - v)|$. Then, by simple algebra, $|v' \cdot (y - v)| \leq \frac{|\lambda|}{2} \|v'\|^2$, and if we let $\lambda \to 0$ we obtain $v' \cdot (y - v) = 0$. Q.E.D.

A good source for more information on this sort of thing is *Basic Classes of Linear Operators* by Gohberg, Goldberg, and Kaashoek (Birkhauser, 2003).

[1]This glosses over one tricky point: how do we know that there is a "closest" point to y in V, i.e. that $\inf_{v \in V} \|y - v\|^2$ is actually attained for some v? To have this, we must require that V be a *closed* subspace. In practice, unless you are very perverse, any subspace you are likely to work with will be closed.