18.06 Problem Set 5 Solution

Due Wednesday, 18 March 2009 at 4 pm in 2-106.
Total: 130 points.

Problem 1: Write down three equations for the line $b = C + Dt$ to go through $b = 7$ at $t = -1$, $b = 7$ at $t = 1$, and $b = 21$ at $t = 2$. Find the least-squares solution $\hat{\vec{x}} = (C, D)^{\mathrm{T}}$. Sketch these three points and the line you found (or use a plotting program).

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

The equations for the line $b = C + Dt$ is

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 7 \\ 7 \\ 21 \end{pmatrix}.$$

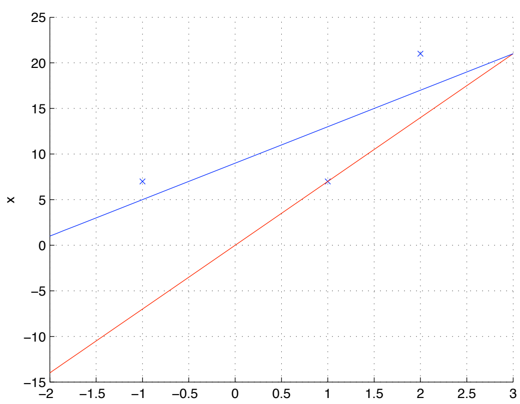
Thus, the least-squares solution is given by solving

$$\begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 7 \\ 7 \\ 21 \end{pmatrix},$$

$$\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 35 \\ 42 \end{pmatrix}$$

Hence, $C = 9, D = 4$ and then $\hat{\vec{x}} = (9\ 4)^{\mathrm{T}}$.

We plot the three points and the lines using MATLAB as follows, where the blue line is the line in the first problem and the red line is the one that passes origin in Problem 2.



```
>> plot(-1, 7, 'x'); plot(1, 7, 'x'); plot(2, 21, 'x');
>> hold on
>> fplot(@(x) 4*x + 9, [-2, 3], 'b');
>> fplot(@(x) 7*x, [-2, 3], 'r');
>> xlabel('t'); ylabel('x');
>> grid
```

Problem 2: For the same three points as in the previous problem, find the best-fit (least-squares) line *through the origin*. (What is the equation of a line through the origin? How many unknown parameters do you have?) Sketch this line on your plot from the previous problem.

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

The equation for the line becomes $b = Dt$ as the line goes through the origin. So we need to find the least-squares solution to the following linear system.

$$\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} (D) = \begin{pmatrix} 7 \\ 7 \\ 21 \end{pmatrix}$$

Similar to Problem 1, we need to solve

$$\begin{pmatrix} -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} (D) = \begin{pmatrix} -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 7 \\ 7 \\ 21 \end{pmatrix}$$

Hence, $6D = 42$, and we have $D = 7$.

Problem 3: If we solve $A^{\mathrm{T}}A\hat{\vec{x}} = A^{\mathrm{T}}\vec{b}$, which of the four subspaces of A contains the error vector $\vec{e} = \vec{b} - A\hat{\vec{x}}$? Which contains the projection $\vec{p} = A\hat{\vec{x}}$? Can $\hat{\vec{x}}$ be chosen to lie completely inside any of the subspaces, and if so which one?

$\boxed{\text{Solution}}$ (15 points = 5+5+5)

Since the error vector is orthogonal to the column space $C(A)$, it lies in the left nullspace $N(A^{\mathrm{T}})$.

The projection $\vec{p}$ is on the column space $C(A)$ (because it is the projection onto the column space).

Since the row space $C(A^{\mathrm{T}})$ and the nullspace $N(A)$ spans the whole space, and we can always modify the vector $\hat{\vec{x}}$ by a vector in $N(A)$ (which does not affect the projection $A\hat{\vec{x}}$). Hence, we can choose $\hat{\vec{x}}$ to be in the row space $C(A^{\mathrm{T}})$.

REMARK: We actually see this phenomenon in Problem 11(d) of Pset 4.

Problem 4: In this problem, you will use 18.02-level calculus to understand why the solution to $A^{\mathrm{T}}A\vec{x} = A^{\mathrm{T}}\vec{b}$ minimizes $\|A\vec{x} - \vec{b}\|$ over all $\vec{x}$, for any arbitrary $m \times n$ matrix A. Consider the function:

$$\begin{aligned} f(\vec{x}) &= \|A\vec{x} - b\|^2 = (A\vec{x} - \vec{b})^{\mathrm{T}}(A\vec{x} - \vec{b}) && (1) \\ &= \vec{x}^{\mathrm{T}}A^{\mathrm{T}}A\vec{x} - \vec{b}^{\mathrm{T}}A\vec{x} - \vec{x}^{\mathrm{T}}A^{\mathrm{T}}\vec{b} + \vec{b}^{\mathrm{T}}\vec{b} && (2) \\ &= \sum_{i,j} B_{ij}x_i x_j - 2\sum_{i,j} A_{ij}b_i x_j + \vec{b}^{\mathrm{T}}\vec{b}, && (3) \end{aligned}$$

where $B = A^{\mathrm{T}}A$. Compute the partial derivatives $\partial f/\partial x_k$ (for any $k = 1, \ldots, n$), and show that $\partial f/\partial x_k = 0$ (true at the minimum of f) leads to the system of n equations $A^{\mathrm{T}}A\vec{x} = A^{\mathrm{T}}\vec{b}$.[1]

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

Fix $k = 1, \ldots, n$. From equation (3), we have

$$\frac{\partial f}{\partial x_k}(\vec{x}) = \sum_j B_{kj}x_j + \sum_i B_{ik}x_i - 2\sum_i A_{ik}b_i.$$

Note that we can simply replace symbols $\sum_j B_{kj}x_j = \sum_i B_{ki}x_i$. Moreover, since B is symmetric, $(B = A^{\mathrm{T}}A = (A^{\mathrm{T}}A)^{\mathrm{T}} = B^{\mathrm{T}})$, we have $\sum_i B_{ki}x_i = \sum_i B_{ik}x_i$. Hence,

$$\frac{\partial f}{\partial x_k}(\vec{x}) = 2\sum_i B_{ik}x_i - 2\sum_i A_{ik}b_i.$$

For $f(\vec{x})$ to achieve its minimum, we need to require that $\frac{\partial f}{\partial x_k}(\vec{x}) = 0$. For this, we need to ask

$$\sum_i B_{ik}x_i = \sum_i A_{ik}b_i \quad \text{for all } k.$$

In matrix notation, this is just the equation for the k-th row of $A^T Ax = A^T b$, recalling that $B = A^T A$. Hence, we have $A^{\mathrm{T}}A\vec{x} = A^{\mathrm{T}}\vec{b}$.

[1] Strictly speaking, by setting $\partial f/\partial x_k = 0$ we are only sure we have an extremum, not a minimum. A little more care is required to establish that it is a minimum—you are *not* required to show this! Informally, $\|A\vec{x} - \vec{b}\|^2$ is clearly increasing if we make $\vec{x}$ arbitrarily large in any direction, so if there is one extremum it can only be a minimum, not a maximum or saddle point (the function is concave-up). A more formal treatment involves the concept of *positive definiteness*, which we will study later in 18.06.

Problem 5: *Matlab problem:* in this problem, you will use Matlab to help you find and plot least-square fits of the function $f(t) = 1/(1 + 25t^2)$ for the $m = 9$ points $t = -0.9, -0.675, \ldots, 0.9$ to a line, quadratic, and higher-order polynomials. First, define the vector $\vec{t}$ of nine t values and the vector $\vec{b}$ of $f(t)$ values, and plot the points as red circles:

```
m = 9
t = linspace(-0.9, 0.9, m)'
b = 1 ./ (1 + 25 * t.^2)
plot(t, b, 'ro')
hold on
```

The command "`hold on`" means that subsequent plots will go on top of this one (normally each time you run `plot` the new plot replaces the old one). To fit to a line $C + Dt$, as in class, we form the $m \times 2$ matrix A whose first column is all ones and whose second column is $\vec{t}$:

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

Now, we have to solve the normal equations $A^{\mathrm{T}}A\vec{x} = A^{\mathrm{T}}\vec{b}$ to find the least-square fit $\vec{x} = (C, D)^{\mathrm{T}}$. In Matlab, however, you can do this with the backslash command, exactly as if you were solving $A\vec{x} = \vec{b}$: Matlab notices that the problem is not exactly solvable and does the least-square solution automatically. Here, we find the least-square line fit and plot it as a blue line for many t values [noting that `x(1)` $= x_1 = C$ and `x(2)` $= x_2 = D$]:

```
x = A \ b
tvals = linspace(-1,1,1000);
plot(tvals, x(1) + x(2) * tvals, 'b-')
```

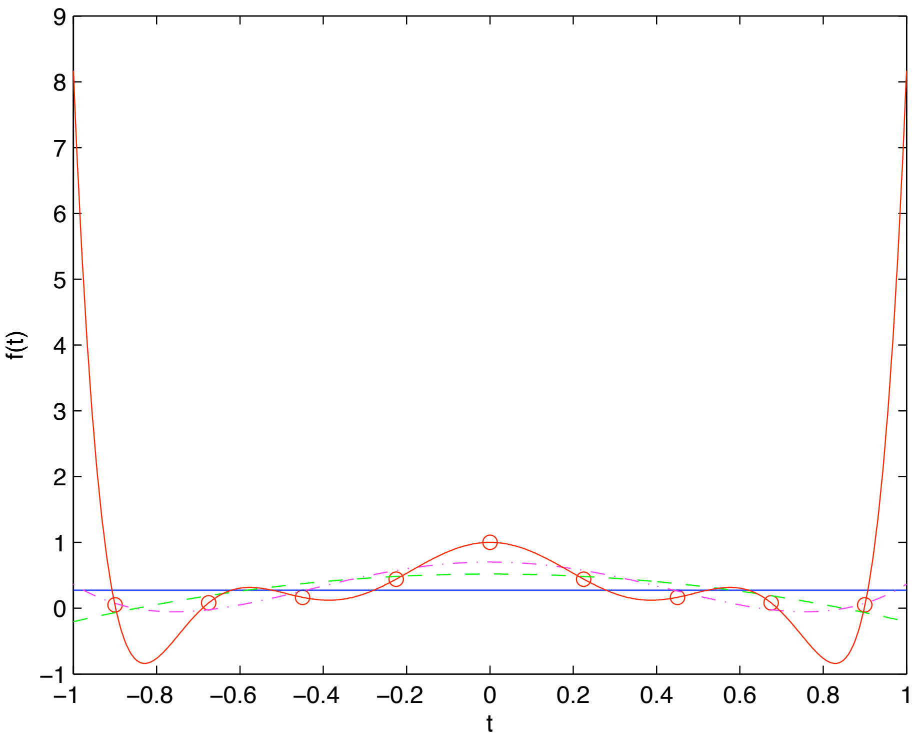
Next, try a quadratic fit, to $C + Dt + Et^2$, plotted as a dashed green line:

```
A = [ones(m,1), t, t.^2]
x = A \ b
plot(tvals, x(1) + x(2) * tvals + x(3) * tvals.^2, 'g--')
```

Finally (figure out the code yourself), fit to a quartic polynomial $C+Dt+Et^2+Ft^3+Gt^4$, and then to a degree-8 polynomial $C+Dt+Et^2+Ft^3+Gt^4+Ht^5+It^6+Jt^7+Kt^8$. Plot your fits, as above. (Turn in a printout of your plots and your code, and the fit coefficients $\vec{x}$ for all four fits.)

Is your fit staying close to the original function $f(t)$? It can be unreliable to try to fit to a high-degree polynomial, due to something called a *Runge phenomenon*.[2]

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




```
>> m = 9;
>> t = linspace(-0.9, 0.9, m)'

t =

  -0.9000
  -0.6750
  -0.4500
  -0.2250
        0
```

[2]Given enough fit parameters, you can fit anything, but such "over-fitting" usually doesn't give useful results. A famous quote attributed by Fermi to von Neumann goes: "With four parameters, I can fit an elephant, and with five I can make him wiggle his trunk."

```
   0.2250
   0.4500
   0.6750
   0.9000

>> b = 1 ./ (1 + 25 * t.^2)

b =

   0.0471
   0.0807
   0.1649
   0.4414
   1.0000
   0.4414
   0.1649
   0.0807
   0.0471

>> plot(t, b, 'ro')
>> hold on
>> A = [ones(m, 1), t]

A =

   1.0000   -0.9000
   1.0000   -0.6750
   1.0000   -0.4500
   1.0000   -0.2250
   1.0000         0
   1.0000    0.2250
   1.0000    0.4500
   1.0000    0.6750
   1.0000    0.9000

>> x = A \ b

x =
```

```
   0.2742
   0.0000

>> tvals = linspace(-1, 1, 1000);
>> plot(tvals, x(1) + x(2) * tvals, 'b-')
>> hold on
>> A = [ones(m,1), t, t.^2]

A =

   1.0000   -0.9000    0.8100
   1.0000   -0.6750    0.4556
   1.0000   -0.4500    0.2025
   1.0000   -0.2250    0.0506
   1.0000         0         0
   1.0000    0.2250    0.0506
   1.0000    0.4500    0.2025
   1.0000    0.6750    0.4556
   1.0000    0.9000    0.8100

>> x = A \ b

x =

   0.5187
   0.0000
  -0.7243

>> plot(tvals, x(1) + x(2) * tvals + x(3) * tvals.^2, 'g--')
>> A = [ones(m,1), t, t.^2, t.^3, t.^4]

A =

   1.0000   -0.9000    0.8100   -0.7290    0.6561
   1.0000   -0.6750    0.4556   -0.3075    0.2076
   1.0000   -0.4500    0.2025   -0.0911    0.0410
   1.0000   -0.2250    0.0506   -0.0114    0.0026
   1.0000         0         0         0         0
   1.0000    0.2250    0.0506    0.0114    0.0026
```

```
   1.0000    0.4500    0.2025    0.0911    0.0410
   1.0000    0.6750    0.4556    0.3075    0.2076
   1.0000    0.9000    0.8100    0.7290    0.6561
>> x = A \ b

x =

   0.7007
  -0.0000
  -2.6385
   0.0000
   2.3015

>> plot(tvals, x(1) + x(2) * tvals + x(3) * tvals.^2 + x(4) * tvals.^3
 + x(5)*tvals.^4, 'm-.')
>> A = [ones(m,1), t, t.^2, t.^3, t.^4, t.^5, t.^6, t.^7, t.^8]

A =

   1.0000   -0.9000    0.8100   -0.7290    0.6561   -0.5905    0.5314
 -0.4783    0.4305
   1.0000   -0.6750    0.4556   -0.3075    0.2076   -0.1401    0.0946
 -0.0638    0.0431
   1.0000   -0.4500    0.2025   -0.0911    0.0410   -0.0185    0.0083
 -0.0037    0.0017
   1.0000   -0.2250    0.0506   -0.0114    0.0026   -0.0006    0.0001
 -0.0000    0.0000
   1.0000         0         0         0         0         0         0
      0         0
   1.0000    0.2250    0.0506    0.0114    0.0026    0.0006    0.0001
  0.0000    0.0000
   1.0000    0.4500    0.2025    0.0911    0.0410    0.0185    0.0083
  0.0037    0.0017
   1.0000    0.6750    0.4556    0.3075    0.2076    0.1401    0.0946
  0.0638    0.0431
   1.0000    0.9000    0.8100    0.7290    0.6561    0.5905    0.5314
  0.4783    0.4305

>> x = A \ b
```

```
x =

   1.0000
  -0.0000
 -14.7838
        0
  82.3065
  -0.0000
 -168.3622
   0.0000
 108.0110
>> plot(tvals, x(1) + x(2) * tvals + x(3) * tvals.^2 + x(4) * tvals.^3
 + x(5) * tvals.^4 + x(6) * tvals.^5 + x(7) * tvals.^6 + x(8) * tvals.^7
 + x(9) * tvals.^8, 'r-')
>> xlabel('t'); ylabel('f(t)');
```

As we fit to a higher and higher degree polynomials, the fit becomes closer and closer to the prescribed points. Since there are 9 points, a degree-8 polynomial (which has 9 coefficients) goes through all of the points exactly. However, in *between* the points the fit polynomial oscillates more and more wildly as the degree is increased. This is a well-known problem when fitting high-degree polynomials to curves, especially at equally-spaced t values. The solution to this problem is outside the scope of 18.06, but lies in the field of *approximation theory*. [It involves choosing the t coordinates more carefully: not to be equally spaced, but to follow a pattern like $t = \cos(n\pi/N)$ for $n = 0, \ldots, N$—this is something called Chebyshev approximation.] The basic point to remember is that using too many fit parameters can lead to very poor results—just because you *can* fit to a function with zillions of parameters doesn't mean that you *should*.

Problem 6: If A has 4 orthogonal columns with lengths 1, 2, 3, and 4, respectively, what is $A^\mathrm{T}A$?

$\boxed{\text{Solution}}$ (5 points)

The matrix A looks like $(\vec{q}_1\ \vec{q}_2\ \vec{q}_3\ \vec{q}_4)$. Then

$$A^\mathrm{T}A = \begin{pmatrix} \vec{q}_1\cdot\vec{q}_1 & \vec{q}_1\cdot\vec{q}_2 & \vec{q}_1\cdot\vec{q}_3 & \vec{q}_1\cdot\vec{q}_4 \\ \vec{q}_2\cdot\vec{q}_1 & \vec{q}_2\cdot\vec{q}_2 & \vec{q}_2\cdot\vec{q}_3 & \vec{q}_2\cdot\vec{q}_4 \\ \vec{q}_3\cdot\vec{q}_1 & \vec{q}_3\cdot\vec{q}_2 & \vec{q}_3\cdot\vec{q}_3 & \vec{q}_3\cdot\vec{q}_4 \\ \vec{q}_4\cdot\vec{q}_1 & \vec{q}_4\cdot\vec{q}_2 & \vec{q}_4\cdot\vec{q}_3 & \vec{q}_4\cdot\vec{q}_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 16 \end{pmatrix}.$$

Problem 7: Give an example of:

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

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

(c) An orthonormal basis for $\mathbb{R}^3$, one vector of which is $\vec{q}_1 = (1,2,3)^{\mathrm{T}}/\sqrt{14}$.

$\boxed{\text{Solution}}$ (15 points = 5+5+5)

(a) Assume that Q has orthonormal columns, then $QQ^{\mathrm{T}} = I$ if and only if Q is **invertible**; this is also equivalent to Q is a square matrix. A counterexample of (a) would be for example,

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, QQ^{\mathrm{T}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

REMARK: QQ^{T} is the projection matrix onto $C(Q)$, which here is the space of vectors $\begin{pmatrix} a & b & 0 \end{pmatrix}^{\mathrm{T}}$.

(b) If one of the vector is 0, then they are both orthogonal and linearly dependent.

(c) We use Gram-Schmidt, to $\vec{q}_1, \vec{a}_2 = (-2,1,0)^{\mathrm{T}}, \vec{a}_3 = (0,-3,2)^{\mathrm{T}}$. (Note that $\vec{q}_1, \vec{a}_2, \vec{a}_3$ are linearly independent.)

We chose $\vec{q}_1^{\mathrm{T}}\vec{a}_2 = 0$, we may take

$$\vec{q}_2 = \vec{a}_2/\|\vec{a}_2\| = (-\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}, 0)^{\mathrm{T}}.$$

Also, we note that we chose $\vec{q}_1^{\mathrm{T}}\vec{a}_3 = 0$, the Gram-Schmidt gives

$$\vec{q}_3 = \frac{\vec{a}_3 - (\vec{a}_3 \cdot \vec{q}_2)\vec{q}_2}{\|\vec{a}_3 - (\vec{a}_3 \cdot \vec{q}_2)\vec{q}_2\|} = \frac{\left(-\frac{6}{5}, -\frac{12}{5}, 2\right)^{\mathrm{T}}}{\left\|\left(-\frac{6}{5}, -\frac{12}{5}, 2\right)^{\mathrm{T}}\right\|} = (-3,-6,5)^{\mathrm{T}}/\sqrt{70}.$$

REMARK: We may choose $\vec{a}_2, \vec{a}_3$ to be any pair of vectors that, along with q_1, are linearly independent. (For example, $\vec{a}_2 = (1,0,0)^{\mathrm{T}}, \vec{a}_3 = (0,1,0)^{\mathrm{T}}$.) Here, we choose $\vec{a}_2, \vec{a}_3$ as above just to make the computation less complicated.

Problem 8: If Q has n orthonormal columns $\vec{q}_1$, $\vec{q}_2$, $\ldots$, $\vec{q}_n$, what combination $x_1\vec{q}_1 + x_2\vec{q}_2 + \cdots + x_n\vec{q}_n$ is closest to a given vector $\vec{b}$? That is, give an explicit expression for $\vec{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$.

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

Since Q has orthonormal columns, $Q^{\mathrm{T}}Q = I$.

To find $\vec{x}$ that gives the combination $x_1\vec{q}_1 + \cdots + x_n\vec{q}_n$ closest to the given $\vec{b}$, we look for the least-squares solution.

$$\vec{x} = (Q^{\mathrm{T}}Q)^{-1}Q^{\mathrm{T}}\vec{b} = Q^{\mathrm{T}}\vec{b} = \begin{pmatrix} \vec{q}_1 \cdot \vec{b} \\ \vdots \\ \vec{q}_n \cdot \vec{b} \end{pmatrix}$$

Problem 9: Find the QR factorization and an orthonormal basis of the column space for the matrix:

$$A = \begin{pmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{pmatrix}.$$

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

We use Gram-Schmidt algorithm.

$$\begin{aligned}
\vec{q}_1 &= \frac{(1,1,1,1)^{\mathrm{T}}}{\|(1,1,1,1)^{\mathrm{T}}\|} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)^{\mathrm{T}}; \\
\vec{q}_2 &= \frac{(-2,0,1,3)^{\mathrm{T}} - \left((\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})(-2,0,1,3)^{\mathrm{T}}\right)(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})^{\mathrm{T}}}{\|(-2,0,1,3)^{\mathrm{T}} - \left((\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})(-2,0,1,3)^{\mathrm{T}}\right)(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})^{\mathrm{T}}\|} \\
&= \left(-\frac{5}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{5}{2}\right)\Big/\sqrt{13}.
\end{aligned}$$

Thus, we write the matrix A as

$$\begin{pmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{5}{2\sqrt{13}} \\ \frac{1}{2} & -\frac{1}{2\sqrt{13}} \\ \frac{1}{2} & \frac{1}{2\sqrt{13}} \\ \frac{1}{2} & \frac{5}{2\sqrt{13}} \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & \sqrt{13} \end{pmatrix}$$

Problem 10: Suppose the QR factorization of $A = QR$ is given by.

$$Q = \begin{pmatrix} 1/\sqrt{3} & 0 & 1/\sqrt{15} \\ 0 & 1/\sqrt{3} & 3/\sqrt{15} \\ 1/\sqrt{3} & 1/\sqrt{3} & -2/\sqrt{15} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{15} \end{pmatrix} \qquad R = \begin{pmatrix} \sqrt{3} & 2\sqrt{3} & -\sqrt{3} \\ 0 & 2\sqrt{3} & \sqrt{3} \\ 0 & 0 & 3\sqrt{15} \end{pmatrix}.$$

Without explicitly computing A, find the least-square solution $\hat{\vec{x}}$ to $A\vec{x} = \vec{b}$ for $\vec{b} = (5, 15, 5, 5)^{\mathrm{T}}$.

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

We first make a derivation copied from the class. We need to solve

$$\begin{aligned} A^{\mathrm{T}}A\hat{\vec{x}} &= A^{\mathrm{T}}\vec{b} \\ (QR)^{\mathrm{T}}(QR)\hat{\vec{x}} &= (QR)^{\mathrm{T}}\vec{b} && (A = QR) \\ R^{\mathrm{T}}Q^{\mathrm{T}}QR\hat{\vec{x}} &= R^{\mathrm{T}}Q^{\mathrm{T}}\vec{b} \\ R^{\mathrm{T}}R\hat{\vec{x}} &= R^{\mathrm{T}}Q^{\mathrm{T}}\vec{b} && Q \text{ is orthogonal, i.e. } Q^{\mathrm{T}}Q = I \\ R\hat{\vec{x}} &= Q^{\mathrm{T}}\vec{b} && R \text{ is invertible.} \end{aligned}$$

So, it suffices to solve $R\hat{\vec{x}} = Q^{\mathrm{T}}\vec{b}$, i.e.

$$\begin{pmatrix} \sqrt{3} & 2\sqrt{3} & -\sqrt{3} \\ 0 & 2\sqrt{3} & \sqrt{3} \\ 0 & 0 & 3\sqrt{15} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{3} & 0 & 1/\sqrt{3} & 1/\sqrt{3} \\ 0 & 1/\sqrt{3} & 1/\sqrt{3} & -1/\sqrt{3} \\ 1/\sqrt{15} & 3/\sqrt{15} & -2/\sqrt{15} & 1/\sqrt{15} \end{pmatrix} \begin{pmatrix} 5 \\ 15 \\ 5 \\ 5 \end{pmatrix}$$

We get $\hat{\vec{x}} = (2, 2, 1)^{\mathrm{T}}$.

Problem 11: Recall that we can define the "length" $\|f(x)\|$ of a function $f(x)$ by $\|f(x)\|^2 = f(x) \cdot f(x)$, where the "dot product" of two functions is $f(x) \cdot g(x) = \int_0^{2\pi} f(x)g(x)\,dx$. With this dot product, three orthonormal functions are $q_1(x) = \sin(x)/\sqrt{\pi}$, $q_2(x) = \sin(2x)/\sqrt{\pi}$, and $q_3(x) = \sin(3x)/\sqrt{\pi}$. If $b(x) = x$, find the closest function $p(x)$ to $b(x)$ (minimizing $\|b(x) - p(x)\|$) in the subspace spanned by q_1, q_2, and q_3. Hint: think about what you would do if these were column vectors rather than functions.

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

We need to find the "orthogonal projection" of $b(x)$ onto the space spanned by q_1, q_2, and q_3, which would minimize $\|b(x) - p(x)\|$. The projection is exactly given by

$$p(x) = \big(b(x)\cdot q_1(x)\big)q_1(x) + \big(b(x)\cdot q_2(x)\big)q_2(x) + \big(b(x)\cdot q_3(x)\big)q_3(x).$$

By calculation, we have

$$\begin{aligned}
b(x)\cdot q_1(x) &= \int_0^{2\pi} \frac{x\sin x}{\sqrt{\pi}}dx = -\frac{x\cos x}{\sqrt{\pi}}\Big|_0^{2\pi} + \int_0^{2\pi}\frac{\cos x}{\sqrt{\pi}}dx = -2\sqrt{\pi},\\
b(x)\cdot q_2(x) &= \int_0^{2\pi} \frac{x\sin 2x}{\sqrt{\pi}}dx = -\frac{x\cos 2x}{2\sqrt{\pi}}\Big|_0^{2\pi} + \int_0^{2\pi}\frac{\cos 2x}{2\sqrt{\pi}}dx = -\sqrt{\pi},\\
b(x)\cdot q_3(x) &= \int_0^{2\pi} \frac{x\sin 3x}{\sqrt{\pi}}dx = -\frac{x\cos 3x}{3\sqrt{\pi}}\Big|_0^{2\pi} + \int_0^{2\pi}\frac{\cos 3x}{3\sqrt{\pi}}dx = -\frac{2}{3}\sqrt{\pi},
\end{aligned}$$

Hence, the projection is

$$p(x) = -2\sin x - \sin 2x - \frac{2}{3}\sin 3x.$$

Problem 12: Define the "dot product" of two functions as

$$f(x)\cdot g(x) = \int_0^\infty f(x)g(x)e^{-x}\,dx.$$

With respect to this dot product, find an orthonormal basis for the subspace of functions spanned by 1, x, and x^2 (i.e. the polynomials of degree 2 or less), using the Gram–Schmidt procedure.

Solution (15 points)

First of all, we include a proof of an 18.01 problem that for any $n \in \mathbb{N}$, $\int_0^\infty x^n e^{-x}dx = n!$. For example, when $n = 1$, we have

$$\int_0^\infty xe^{-x}dx = -xe^{-x}\big|_0^\infty + \int_0^\infty e^{-x}dx = 1.$$

The same integration by part technique can be used to prove the formulas in general. If we have prove the integral identity for some n, then for $n+1$ we have

$$\int_0^\infty x^{n+1}e^{-x}dx = -x^{n+1}e^{-x}\big|_0^\infty + \int_0^\infty (n+1)x^n e^{-x}dx = (n+1)\cdot n! = (n+1)!.$$

Now we return to the problem and start with the function 1.

$$1 \cdot 1 = \int_0^\infty e^{-x} dx = 1.$$

Hence the function 1 is an normal basis of the space spanned by $1, x, x^2$ under the prescribed norm; use q_1 to denote the constant function 1.

Since $q_1(x) \cdot x = \int_0^\infty x e^{-x} dx = 1$, we know that $x - 1 \cdot q_1(x) = x - 1$ is orthogonal to the function $q_1(x) = 1$. We calculate the norm of $x - 1$ as follows

$$\|x-1\|^2 = \int_0^\infty (x-1)^2 e^{-x} dx = \int_0^\infty \left(x^2 e^{-x} - 2x e^{-x} + e^{-x}\right) dx = 2! - 2 \times 1 + 1 = 1.$$

Hence, we may take $q_2 = x - 1$ as the second vector of the orthonormal basis of the space.

Now, we calculate $q_1 \cdot x^2 = \int_0^\infty x^2 e^{-x} dx = 2$ and $q_2 \cdot x^2 = \int_0^\infty x^2 (x-1) e^{-x} dx = 3! - 2! = 4$. Hence $x^2 - 4q_2(x) - 2q_1(x) = x^2 - 4x + 2$ is orthogonal to the functions $q_1(x)$ and $q_2(x)$.

$$\begin{aligned} \|x^2 - 4x + 2\|^2 &= \int_0^\infty (x^2 - 4x + 2)^2 e^{-x} dx = \int_0^\infty \left(x^4 - 8x^3 + 20x^2 - 16x + 4\right) dx \\ &= 4! - 8 \times 3! + 20 \times 2! - 16 \times 1 + 4 = 4. \end{aligned}$$

Hence, we should take $q_3(x) = \frac{1}{2}x^2 - 2x + 1$.

REMARK: the polynomials that we obtain by this process (with a slightly different sign convention) are called the *Laguerre polynomials*; you can find out a lot more about these famous polynomials by googling this name.