18.06 Problem Set 5

Due Wednesday, 18 March 2008 at 4pm in 2-106.

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 $\hat{\mathbf{x}} = (C, D)^T$. Sketch these three points and the line you found (or use a plotting program).

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.

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

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

$$f(\mathbf{x}) = \|A\mathbf{x} - b\|^2 = (A\mathbf{x} - \mathbf{b})^T(A\mathbf{x} - \mathbf{b}) \tag{1}$$
$$= \mathbf{x}^T A^T A\mathbf{x} - \mathbf{b}^T A\mathbf{x} - \mathbf{x}^T A^T\mathbf{b} + \mathbf{b}^T\mathbf{b} \tag{2}$$
$$= \sum_{i,j} B_{ij}x_i x_j - 2\sum_{i,j} A_{ij}b_i x_j + \mathbf{b}^T\mathbf{b}, \tag{3}$$

   where $B = A^T A$. 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 $A^T A\mathbf{x} = A^T\mathbf{b}$.[1]

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 $\mathbf{t}$ of nine t values and the vector $\mathbf{b}$ 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 $\mathbf{t}$:

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

   Now, we have to solve the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$ to find the least-square fit $\mathbf{x} = (C, D)^T$. In Matlab, however, you can do this with the backslash command, exactly as if you were solving $A\mathbf{x} = \mathbf{b}$: 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)` $= x_2 = D$]:

[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, $\|A\mathbf{x} - \mathbf{b}\|^2$ is clearly increasing if we make $\mathbf{x}$ 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.

```
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 **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]

6. If A has 4 orthogonal columns with lengths 1, 2, 3, and 4, respectively, what is A^TA?

7. Give an example of:
   (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 $\mathbb{R}^3$, one vector of which is $\mathbf{q}_1 = (1,2,3)^T/\sqrt{14}$.

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

9. Find the QR factorization and an orthonormal basis of the column space for the matrix:
$$A = \begin{pmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 3 \end{pmatrix}.$$

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 $\hat{\mathbf{x}}$ to $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} = (5,15,5,5)^T$.

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.

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.

[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."