HW 4

Due by lecture on Weds, Mar 7

Remember that you may (and should!) talk about the problems amongst yourselves, or discuss them with me or the TA, providing attribution for any good ideas you might get – but your final write-up should be your own.

1: A different sort of normal If M is a symmetric and positive definite matrix, we can define the M inner product:

$$\langle x, y \rangle_M = x^T M y.$$

Similarly, we can define the M-norm: $\|x\|_M = \sqrt{x^T M x}$.

1. Derive a version of the normal equations for minimizing

$$\phi_M(x) = \|Ax - b\|_M^2$$

2. Suppose $M = LL^T$ is a Cholesky factorization, where L is lower triangular. Show that for any z,

$$\|z\|_M^2 = \|L^T z\|_2^2$$

3. Fill in the following MATLAB code fragment to solve the M-norm least squares problem

```
LT = chol(M);             % Returns the upper triangular Cholesky factor
[Q,R] = qr( ... , 0);     % Compute an economy QR decomposition
x = R\(Q'*( ... ));       % Solve the M-norm LS problem
```

2: SVD stuff Let $A \in \mathbb{R}^{m \times n}$, $m > n$, and let $A = U_1 \Sigma_{11} V^T$ be the economy SVD.

1. Show that $\|U_1 x\|_2 = \|x\|_2$. *Hint:* Remember that because U is orthogonal, $\|U^T y\| = \|y\|$ for any y.

2. Show that $\|Av_1\|_2 = \sigma_1 \|v_1\|_2$, where v_1 is the first column of V.

3. Show that for any x, $\|Ax\|_2 \leq \sigma_1 \|x\|$. Together with the previous observation, this tells us that
$$\|A\|_2 = \max_{\|x\|=1} \|Ax\|_2 = \sigma_1.$$

4. Show that
$$\|A\|_F^2 = \sum_j \sigma_j^2.$$
Note: it may help to write the squared Frobenius norm of A as the sum of squared Euclidean norms of the columns of A.

3: A fitting end Load the files `hw4A.txt` and `hw4b.txt` for the A matrix and b vector for this assignment:

```
A = load('hw4A.txt');
b = load('hw4b.txt');
```

The *coefficient of determination* R^2 is often used to evaluate how well a linear regression model fits data. Given a vector x of regression coefficients, we can write R^2 in linear algebraic terms as
$$R^2 = 1 - \frac{\|r\|^2}{\|b - \bar{b}e\|^2}$$
where $r = Ax - b$ and $\bar{b}$ is the mean of b. If $R^2 = 1$, we have a perfect fit; if $R^2 = 0$, our model predictions are basically no better than a constant predictor.

1. Use ordinary least squares to compute regression coefficients:
```
x0 = A\b;
```
What is the R^2 score for this fit?

2. Now, do a regression using only the first 10 data points:
```
x1 = A(1:10,:)\b(1:10);
```
What is the R^2 score for this model (computed over all the data)?

3. Use the top two singular vectors of the first 10 data points to compute another fit:

```
[U,S,V] = svd(A(1:10,:), 0);
x = V(:,1:2)*(S(1:2,1:2)\(U(:,1:2)'*b(1:10)));
```

What is the R^2 score for this fit?

4. Do a regression based only on the first two columns of A using ordinary least squares. What is the R^2 score for this fit?

5. Do the above experiments (ordinary least squares, truncated SVD, and least squares on the first two columns) 1000 times, but each time with a randomly chosen set of ten data points (use `randperm` to generate the samples). Plot a histogram of R^2 values for the fits from each of the three methods.