CS 6210: Assignment 5

Due: Wednesday, October 20, 2010 (In Class or in Upson 5153 by 4pm)

Scoring for each problem is on a 0-to-5 scale (5 = complete success, 4 = overlooked a small detail, 3 = good start, 2 = right idea, 1 = germ of the right idea, 0 = missed the point of the problem.) One point will be deducted for insufficiently commented code. Unless otherwise stated, you are expected to utilize fully MATLAB'S vectorizing capability subject to the constraint of being flop-efficient. Test drivers and related material are posted at `http://www.cs.cornell.edu/courses/cs6210/2010fa/`. For each problem submit output and a listing of all scripts/functions that *you* had to write in order to produce the output. You are allowed to discuss *background* issues with other students, but the codes you submit must be your own.

P1. (Low Rank LS)

Consider the least squares problem

$$\min_{x \in \mathbb{R}^n} \| FG^T x - b \|_2$$

where $F \in \mathbb{R}^{m\times r}$, $G \in \mathbb{R}^{n\times r}$, and $b \in \mathbb{R}^n$. Assume that $\text{rank}(F) = \text{rank}(G) = r < n \le m$. Write a MATLAB function `xB = LowRankLS(F,G,b)` that returns the basic solution. This means that you must compute the qr-with-column-pivoting factorization of $A = FG^T$. In reasoning about the efficiency of your method, assume $r << n$. Make effective use of the MATLAB function `qr`. Test your implementation with `P1` and submit output.

P2. (Translation LS)

Suppose $A, B \in \mathbb{R}^{m\times 3}$ are data matrices that result when taking an "xyz-snapshot" of protein molecules $\mathcal{A}$ and $\mathcal{B}$ respectively. In particular, assume that the $A(i,:)$ and $B(i,:)$ specify the centers of the i-th amino acid in $\mathcal{A}$ and $\mathcal{B}$. We want to know if we can translate and rotate $\mathcal{B}$ so that it coincides with $\mathcal{A}$ thereby shedding light on the question "Are we looking at the same protein?"

If we translate $\mathcal{B}$ by $v \in \mathbb{R}^3$ then this corresponds to adding v^T to each row in B. I.e., $B \to B + ev^T$ where $e =$ `ones(m,1)`. If we rotate the translated $\mathcal{B}$ by a 3-by-3 orthogonal matrix Q and use the Frobenius norm as a metric, then we are interested in minimizing

$$\phi(Q, v) = \| A - (B + ev^T)Q \|_F.$$

Complete the following function so that it performs as specified

```
  function [Q,v] = MinPhi(A,B)
% A and B are m-by-n matrices with n<=m.
% Q is an n-by-n orthogonal matrix and v is a column n-vector with the
% property that norm( (A - (B + e*v')Q,'fro' ) is minimized.
```

Test your implementation on the script `P2` and submit output.

A note about translation. The minimizer of the full rank least squares problem $\min\| Fx - b \|$ is the solution to the normal equations $F^T F x = F^T b$. You can use this fact to prove that if $A_1, B_1 \in \mathbb{R}^{m\times n}$ then the vector $v \in \mathbb{R}^n$ that minimizes $\| A_1 - (B_1 + ev^T) \|_F$ where $e =$ ones(m,1) is given by

$$v_{opt} = \frac{1}{m}(A_1 - B_1)^T e.$$

A note about rotation. If $A, B \in \mathbb{R}^{m\times n}$ and $Q \in \mathbb{R}^{n\times n}$ is orthogonal then

$$\| A - BQ \|_F^2 \;=\; \| A \|_F^2 + \| B \|_F^2 - 2\text{tr}(A^T BQ)$$

where $\text{tr}(C)$ = `sum(diag(C))` = the sum of C's eigenvalues. To find the minimizing Q we note that if $A^TB = U\Sigma V^T$ is the SVD of A^TB, then we need to find that Q which maximizes

$$\text{tr}(A^TBQ) = \text{tr}(U^T(U\Sigma V^TQ)U) = \text{tr}(U\Sigma V^TQ) = \text{tr}(\Sigma V^TQU) = \sum_{i=1}^{n} \sigma_i z_{ii}$$

where $Z = V^TQU$. This is accomplished by setting $Q = VU^T$.

Finally, remember that $\| A \|_F^2 = \sum_i \sum_j a_{ij}^2 \;=\; \text{tr}(A^TA)$.

P3. (Window LS)

Complete the following function so that it performs as specified

```
  function rho = WindowLS(A,b,p)
% A is m-by-n, b is m-by-1, and p satisfies n < p <= m
% rank(A(i:i+p-1,:)) = n for i=1:m-p+1
% rho is a column (m-p+1)-vector where rho(i) is the minimum value of
% norm(A(i:i+p-1,:)x - b(i:i+p-1),2)
```

Hint: Read §12.5 in GVL. Test your implementation with the script `P3` and submit output.

Challenge Problem 5. (The p-TLS Problem)

Suppose $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $n < m$. Consider the following problem where p satisfies $0 \leq p \leq n$

$$\min_{\substack{b + r \in \text{ran}(A+E) \\ E(:,1{:}p) = 0}} \| E \|_F^2 + \| r \|_2^2$$

Think of this as a total least squares problem where the first p columns of A are exact. (If $p = 0$, then we have the traditional TLS problem. If $p = n$, then we have the traditional LS problem.) If E_{pTLS} and r_{pTLS} are optimal, then there is a vector $x_{pTLS} \in \mathbb{R}^n$ such that $(A + E_{pTLS})x_{pTLS} = b + r_{pTLS}$. We refer to $\{x_{pTLS}, E_{pTLS}, r_{pTLS}\}$ as the pTLS solution.

```
    function [x,E,r] = pTLS(A,b,p)
%   A m-by-n, b m-by-1, 0 <=p <= n
%   Assume that A has full rank and that the pTLS problem has a unique solution
%   {x_pTLS, E_pTLS, r_pTLS.} Then x = x_pTLS, E = E_pTLS, and r = r_pTLS
```

Hint: Convert the problem into an equivalent easy-to-solve problem where A is upper triangular. Test your implementation with the script `C5` and submit output..