

Problem 12.2 worth 6 points

Least squares with orthonormal columns. Suppose the $m \times n$ matrix Q has orthonormal columns and b is an m -vector. Show that $\hat{x} = Q^T b$ is the vector that minimizes $\|Qx - b\|^2$. What is the complexity of computing \hat{x} , given Q and b , and how does it compare to the complexity of a general least squares problem with an $m \times n$ coefficient matrix?

Solution: $\hat{x} = Q^T b$ because the full equation is $\hat{x} = (Q^T Q)^{-1} Q^T b$. The reason we eliminate $(Q^T Q)^{-1}$ is because $Q^T Q = I$ (orthonormal rows times orthonormal columns), the inverse of the identity matrix is the identity matrix, and $I Q^T b = Q^T b$.

The complexity of a least square problem $\|Ax - b\|^2$ where A is an $m \times n$ coefficient matrix is $2mn^2$. The complexity of solving for \hat{x} and the complexity of a matrix vector multiplication are the same: $2mn$. It is much more efficient to find $Q^T b$ than solve a general least squares equation.

Problem 12.4 worth 8 points

Weighted least squares. In least squares, the objective (to be minimized) is

$$\|Ax - b\|^2 = \sum_{i=1}^m (\tilde{a}_i^T x - b_i)^2,$$

where \tilde{a}_i^T are the rows of A , and the n -vector x is chosen. In *weighted least squares problem*, we minimize the objective

$$\sum_{i=1}^m w_i (\tilde{a}_i^T x - b_i)^2,$$

where w_i are given possible weights. The weights to the different components of the residual vector. (The objective of the weighted least squares problem is the square of the weighted norm, $\|Ax - b\|_w^2$, as defined in exercise 3.28.

- Show that the weighted least squares objective can be expressed as $\|D(Ax - b)\|^2$ for an appropriate diagonal matrix D . This allows us to solve the weighted least squares problem as a standard least squares problem, by minimizing $\|Bx - d\|^2$, where $B = DA$ and $d = Db$.
- Show that when A has linearly independent columns, so does the matrix B .
- The least squares approximate solution is given by $\hat{x} = (A^T A)^{-1} A^T b$. Give a similar formula for the solution of the weighted least squares problem. You might want to use the matrix $W = \text{diag}(w)$ in your formula.

Solution:

- First we will distribute the weight attribute into the summation. When we distribute a number into a squared problem we have to take the square root of it for it to balance.

$$\begin{aligned} \sum_{i=1}^m w_i (\tilde{a}_i^T x - b_i)^2 \\ \sum_{i=1}^m (\sqrt{w_i} (\tilde{a}_i^T x - b_i))^2 \end{aligned}$$

Since $\tilde{a}_i^T x$ will produce a scalar, we don't want to multiply it multiple times by the weight, then sum all those numbers. So our matrix D should have a square root value of the weight index only in the

index that corresponds to the row of A and index of b we are currently looking at.

$$D = \begin{bmatrix} \sqrt{w_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{w_2} & 0 & \cdots & 0 \\ 0 & 0 & \sqrt{w_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \sqrt{w_m} \end{bmatrix}$$

This shows us two things: first, B will be a sort of identity matrix, following the pattern of D it would contain the weighted rows of A as a scalar at each of it's diagonal indices. Second, Db will make the b_i^{th} index equally weighted as well.

- (b) If A has linearly independent columns then so will B . This is shown with D solved in part a, when we multiply the matrix D by the matrix A we would get a matrix whose only indices are diagonal. since only the diagonals are filled then we could not possibly get one column from another, since we will always have 0's matching up with the weighted index. Therefore B would have linearly independent columns (and rows).

(c)

$$\hat{x} = (B^T B)^{-1} B^T d$$

Problem 12.1 worth 3 points

Approximating a vector as a multiple of another one. In the special case $n = 1$, the general least squares problem (12.1) reduces to finding a scalar x that minimizes $\|ax - b\|^2$, where a and b are m -vectors. (We write the matrix A here in lower case, since it is an m -vector). Assuming a and b are non-zero, show that $\|a\hat{x} - b\|^2 = \|b\|^2(\sin \theta)^2$, where $\theta = \angle(a, b)$. This shows that the optimal relative error in approximating one vector by a multiple of another one depends on their angle.

Solution: First, $(\sin \theta)^2 = 1 - (\cos \theta)^2 \longrightarrow \cos \theta = \frac{a^T b}{\|a\| \|b\|} \longrightarrow (\cos \theta)^2 = \frac{(a^T b)^2}{\|a\|^2 \|b\|^2}$

$$\|a\hat{x} - b\|^2 = \|b\|^2(\sin \theta)^2$$

$$\hat{x} = (a^T a)^{-1} a^T b = \frac{a^T b}{\|a\|^2}$$

$$\|a^T \left(\frac{a^T b}{\|a\|^2} \right) - b\|^2 = \|b\|^2(\sin \theta)^2$$

$$\|a^T \left(\frac{a^T b}{\|a\|^2} \right) - b\| \cdot \|a^T \left(\frac{a^T b}{\|a\|^2} \right) - b\| = \|b\|^2(\sin \theta)^2$$

$$\|a\|^2 \frac{(a^T b)^2}{\|a\|^4} - 2 \frac{(a^T b)^2}{\|a\|^2} + \|b\|^2 = \|b\|^2(\sin \theta)^2$$

$$\frac{(a^T b)^2}{\|a\|^2} - 2 \frac{(a^T b)^2}{\|a\|^2} + \|b\|^2 = \|b\|^2(\sin \theta)^2$$

$$-\frac{(a^T b)^2}{\|a\|^2} + \|b\|^2 = \|b\|^2(\sin \theta)^2$$

$$\left(\frac{1}{\|b\|^2} \right) \left(-\frac{(a^T b)^2}{\|a\|^2} + \|b\|^2 \right) = (\sin \theta)^2$$

$$1 - \frac{(a^T b)^2}{\|a\|^2 \|b\|^2} = \sin^2 \theta$$

$$1 - \cos^2 \theta = \sin^2 \theta$$

Problem 12.8a *worth 3 points*

Least squares and QR factorization. Suppose A is an $m \times n$ matrix with linearly independent columns and QR factorization $A = QR$, and b is an m -vector. The vector $A\hat{x}$ is the linear combination of the columns of A that is closest to the vector b , i.e., it is the projection of b onto the set of linear combinations of the columns of A .

1. Show that $A\hat{x} = QQ^T b$. (The matrix QQ^T is called the *projection matrix*.)

Solution: Since R is invertible, we can also solve for $Q = AR^{-1}$

$$\hat{x} = A^\dagger b$$

$$A^\dagger = R^{-1}Q^T$$

$$A\hat{x} = QQ^T b$$

$$AA^\dagger b = QQ^T b$$

$$AR^{-1}Q^T b = QQ^T b$$

$$QQ^T b = QQ^T b$$