# Numerical Analysis

Math 370 Fall 2004 MWF 2:30 - 3:25pm
 Fowler North 5

*Worksheet 21*

**SUMMARY** Using the QR Algorithm to Solve Least-Squares Problems
**READING** Recktenwald, Sec 9.2, pp 473-499

**Coefficient of Determination**
As we complete our discussion of least squares problems it is useful to note that a popular measure of the fit of a curve (or a linear combination of functions) to the given data is called the coefficient of determination and denoted by $R^2$.

$$R^2 = 1 - \frac{m-1}{m-n-1}\frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \bar{y})^2}$$

where $\bar{y}$ is the mean of the $y$ data, and $\hat{y}_i = P(x_i) = \sum_{j=1}^{n} a_j f_j(x_i)$
Most packages or routines which do curve fitting will generate this statistic. It is a measure of the "goodness" of the fit.

For example if you use `linefit2.m` in the fit directory of the NMM toolbox, it produces the $R^2$ statistic.

## QR Algorithm

One way to obtain a solution to an over-determined system of equations is to use the **QR Algorithm** or **QR factorization**. We try and find to matrices $Q$ and $R$ such that $D = QR$. The matrix $Q$ is an orthogonal matrix (i.e. $Q^{-1} = Q^T$) and $R$ is an upper-triangular matrix. We won't discuss how the QR algorithm works, but you can read up on it on page 485 of Recktenwald.

Note the QR Algorithm is so important and fundamental in MATLAB that when one uses the backslash operator in MATLAB what happens is that if the matrix is $m$ by $n$ with $m > n$ MATLAB automatically applies QR factorization to the matrix and finds the least squares vector.
So, we can rewrite

$$Da = y$$

as

$$QRa = y$$

**Example**
5. Using Linear Algebra, we can get an explicit formula for $a$ in terms $R^{-1}$ and $Q^T$ in order to solve $Da = y$.(What MATLAB command would solve this equation?)

**Summary**

The least-squares solution to the overdetermined system $A\vec{c} = \vec{y}$ is equivalent to finding the minimum of $||Q^T\vec{y} - R\vec{c}||_2$ where $A = QR$ is the QR factorization of $A$.

**Exercise**

Suppose we have the data $(1, 1)$, $(2, 2)$, $(4, 2)$ and $(5, 3)$.

6. What is the matrix system are trying to solve (i.e. $Da = y$)?

7. What is $n$? What is $m$? What are the dimensions of $D$?

8. Use the MATLAB command `qr` to find $Q$ and $R$. Do they have the properties you expect?

9. Find the vector $a$ and write down the equation of the line of best fit $y = c_1x + c_2$

10. Find the line of best fit using the MATLAB `linefit` command and graph the data and the line. (Write down what commands you use.)

11. How do your answers from the two methods compare?