Lecture # 7
Modified Gram–Schmidt and Normal Equations in Least Squares Solutions

To compute

$$\begin{aligned} X &= Q_1 R = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \\ Q_1 &= (\mathbf{q}_1, \ldots, \mathbf{q}_n), \quad R = (r_{jk}) \end{aligned}$$

we use the **Modified Gram–Schmidt (MGS) Algorithm**

$r_{11} = \|\mathbf{x}_1\|_2$; $\mathbf{q}_1 = \mathbf{x}_1 / r_{11}$;
for $k = 2{:}\, n$
    $\mathbf{s}_k = \mathbf{x}_k$;
    for $j = 1{:}\, k-1$
        $r_{jk} = \mathbf{q}_j^T \mathbf{s}_k$;
        $\mathbf{s}_k = \mathbf{s}_k - r_{jk}\mathbf{q}_j$;
    end;
    $r_{kk} = \|\mathbf{s}_k\|_2$;
end;

To solve for $\mathbf{y}_{LS}$ from

$$\|\mathbf{b} - X\mathbf{y}_{LS}\|_2 = \min_{\mathbf{y} \in \mathbf{R}^n} \|\mathbf{b} - X\mathbf{y}\|_2$$

treat $\mathbf{b}$ as an extra column of X (you can do this in Householder, too)!

$\mathbf{r} = \mathbf{b}$;
for $j = 1{:}\, n$
    $c_j = \mathbf{q}_j^T \mathbf{r}$;
    $\mathbf{r} = \mathbf{r} - c_j \mathbf{q}_j$;
end;
$r_{LS} = \mathbf{r}$; % You do not need to code this line
$\mathbf{c} = (c_1, \ldots, c_n)^T$;
Solve $R\mathbf{y}_{LS} = \mathbf{c}$;

The solution $\mathbf{y}_{LS}$ and the residual $\mathbf{r}_{LS}$ are as good as Householder.

But, DO NOT DO THIS

$$\begin{aligned} \mathbf{c} &= Q_1^T \mathbf{b} \\ \mathbf{r}_{LS} &= \mathbf{b} - Q_1 \mathbf{c} \\ R\mathbf{y}_{LS} &= \mathbf{c} \end{aligned}$$

It yields a poor solution.

The first procedure obtains a good solution because it is related to Householder Q–R on

$$\begin{array}{c} n \\ m \end{array} \begin{pmatrix} 0 \\ X \end{pmatrix} = \hat{Q} \begin{pmatrix} R \\ 0 \end{pmatrix}.$$

The matrix $\hat{Q}$ is given by

$$\hat{Q} = \hat{H}_1 \cdots \hat{H}_n$$

where

$$\begin{aligned} \hat{H}_k &= I - \mathbf{w}_k \mathbf{w}_k^T, \quad \|\mathbf{w}_k\|_2 = \sqrt{2} \\ \mathbf{w}_k &= \begin{pmatrix} -\mathbf{e}_k \\ \mathbf{q}_k \end{pmatrix}. \end{aligned}$$

Here $\mathbf{q}_k$ is the same vector, both mathematically and numerically, as is produced by modified Gram–Schmidt.

The first procedure uses this structure to solve

$$\| \begin{pmatrix} 0 \\ \mathbf{b} \end{pmatrix} - \begin{pmatrix} 0 \\ X \end{pmatrix} \mathbf{y}_{LS} \|_2 = \min_{\mathbf{y} \in \mathbf{R}^n} \| \begin{pmatrix} 0 \\ \mathbf{b} \end{pmatrix} - \begin{pmatrix} 0 \\ X \end{pmatrix} \mathbf{y} \|_2.$$

This connection was discovered by Charles Sheffield (by accident) in the 1960's.

Cholesky Factorization of the Normal Equations

Note that the residual

$$\mathbf{r}_{LS} = \mathbf{b} - X\mathbf{y}_{LS}$$

satisfies

$$X^T \mathbf{r}_{LS} = 0$$

yields

$$X^T \mathbf{r}_{LS} = X^T(\mathbf{b} - X\mathbf{y}_{LS}) = 0$$

which reorganizes to

$$X^T X \mathbf{y}_{LS} = X^T \mathbf{b}. \tag{1}$$

Equation (1) is call the normal equations. The matrix $A = X^T X$ is clearly symmetric, and also *positive definite*. The latter means that

$$\mathbf{y}^T A \mathbf{y} > 0, \quad \mathbf{y} \neq 0.$$

That follows from

$$\mathbf{y}^T A \mathbf{y} = \mathbf{y}^T X^T X \mathbf{y} = \|X\mathbf{y}\|_2^2 > 0$$

for all $\mathbf{y} \neq 0$ since $\text{rank}(X) = n$.

Thus we can solve (1) using the Cholesky factorization of A, given by

$$A = X^T X = R^T R$$

where $R \in \mathbf{R}^{n \times n}$ is upper triangular with positive diagonal elements. We then solve

$$R^T R \mathbf{y}_{LS} = X^T \mathbf{b}$$

using forward and back substitution.

For review, the Cholesky algorithm is given by (a different algorithm from that given in class, using the MATLAB ability to expand matrices).

function R=my_chol(A)
$R = \sqrt{A(1,1)}$;
for $k = 2n$
  Solve $R'\mathbf{v} = A(1:k-1, k)$;
  $rkk = \sqrt{A(k,k) - \mathbf{v}^T \mathbf{v}}$;
  $R = [R\,\mathbf{v}; zeros(1, k-1)\, rkk]$;
end;
end; *my_chol*

It is no coincidence that we labeled this upper triangular matrix R.

For Q–R factorization, we have

$$X = Q \begin{pmatrix} R \\ 0 \end{pmatrix} = Q_1 R,$$

where Q is orthogonal and Q_1 is left orthogonal. Note that

$$\begin{aligned} A &= X^T X = \begin{pmatrix} R^T & 0 \end{pmatrix} Q^T Q \begin{pmatrix} R \\ 0 \end{pmatrix} = R^T Q_1^T Q_1 R \\ &= R^T R. \end{aligned}$$

Since the R from Cholesky factorization and that from Q–R factorization are both unique, they must be the same R! (Mathematically, of course). Below is an example (I changed it slightly from the one in class).

Example 1 *Let X be the Läuchli matrix*

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & \delta & 0 & 0 \\ 0 & 0 & \delta & 0 \\ 0 & 0 & 0 & \delta \end{pmatrix}$$

where $\delta = sqrt(eps)$,$eps = 2^{-52}$. Householder Q–R factorization yields

$$Q_H = \begin{pmatrix} -1 & 1.05367e-08 & 6.08337e-09 & 4.30159e-09 \\ -1.49012e-08 & -0.707107 & -0.408248 & -0.288675 \\ -0 & 0.707107 & -0.408248 & -0.288675 \\ -0 & 0 & 0.816497 & -0.288675 \\ -0 & 0 & 0 & 0.866025 \end{pmatrix},$$

$$R_H = \begin{pmatrix} -1 & -1 & -1 & -1 \\ 0 & 2.10734e-08 & 1.05367e-08 & 1.05367e-08 \\ 0 & 0 & 1.82501e-08 & 6.08337e-09 \\ 0 & 0 & 0 & 1.72064e-08 \end{pmatrix}$$

Cholesky factorization of the normal equations gives us

$$R = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1.4901e-08 & 0 & 0 \\ 0 & 0 & 1.4901e-08 & 0 \\ 0 & 0 & 0 & 1.4901e-08 \end{pmatrix}$$

We gave this problem the right hand side

$$b = X \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \delta * \begin{pmatrix} -7.4506e-09 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}$$

which should yield

$$\mathbf{y}_{LS} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{r}_{LS} = \delta * \begin{pmatrix} -7.4506e-09 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}$$

Householder Q–R factorization yields the solution $\hat{\mathbf{y}}_{LS}$ *and residual* $\hat{\mathbf{r}}_{LS}$ *such that*

$$\|\hat{\mathbf{y}}_{LS} - \mathbf{y}_{LS}\|_2 / \|\mathbf{y}_{LS}\|_2 = 2.8305e-16, \quad \|\hat{\mathbf{r}}_{LS} - \mathbf{r}_{LS}\|_2 / \|\mathbf{r}_{LS}\|_2 = 5.5511e-16$$

Normal equations gets the solution

$$\hat{\mathbf{y}}_{LS} = \begin{pmatrix} 4 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{r}_{LS} = \begin{pmatrix} 0 \\ -3.7253e-08 \\ 2.2352e-08 \\ 2.2352e-08 \\ 2.2352e-08 \end{pmatrix}$$

Neither $\hat{\mathbf{y}}_{LS}$ *nor* $\hat{\mathbf{r}}_{LS}$ *has any correct digits.*

If we take this example further, say, for the Läuchli matrix

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1e-10 & 0 & 0 & 0 \\ 0 & 1e-10 & 0 & 0 \\ 0 & 0 & 1e-10 & 0 \\ 0 & 0 & 0 & 1e-10 \end{pmatrix}$$

The Householder and Gram–Schmidt Q–R decompositions of X *are fine in floating point arithmetic, but when we compute the normal equations in MATLAB we get*

$$A = \mathrm{f}\ell(X^T X) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

rounded from

$$X^T X = \begin{pmatrix} 1 = 1e-20 & 1 & 1 & 1 \\ 1 & 1+1e-20 & 1 & 1 \\ 1 & 1 & 1+1e-20 & 1 \\ 1 & 1 & 1 & 1+1e-20 \end{pmatrix}.$$

Thus the normal equations algorithm simply fails.