Lecture # 5
The Linear Least Squares Problem

Let $X \in \mathbf{R}^{m\times n}, m \geq n$ be such that $\text{rank}(X) = n$. That is,

$$X\mathbf{y} = 0, \quad \textbf{iff } \mathbf{y} = 0.$$

The problem is to find $\mathbf{y}_{LS}$ such that

$$\|\mathbf{b} - X\mathbf{y}_{LS}\|_2 = \min_{\mathbf{y}\in\mathbf{R}^n} \|\mathbf{b} - X\mathbf{y}\|_2^2 \tag{1}$$

We also want

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

Our approach, compute the Q–R decomposition , that is,

$$X = Q \begin{matrix} n \\ m-n \end{matrix} \begin{pmatrix} R \\ 0 \end{pmatrix},$$

where $Q \in \mathbf{R}^{m\times m}$ is orthogonal and $R \in \mathbf{R}^{n\times n}$ is upper triangular and nonsingular.

That yields a procedure of the form

1. Use $[Q, R] = qr(X)$ to get Q and R.
2. Compute

$$\mathbf{c} = \begin{matrix} n \\ m-n \end{matrix} \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{pmatrix} = Q^T\mathbf{b}.$$

3. Solve

$$R\mathbf{y}_{LS} = \mathbf{c}_1,$$

and compute

$$\mathbf{r}_{LS} = Q \begin{pmatrix} 0 \\ \mathbf{c}_2 \end{pmatrix}.$$

Note that

$$\begin{aligned} X^T\mathbf{r}_{LS} &= \begin{pmatrix} R^T & 0 \end{pmatrix} Q^T Q \begin{pmatrix} 0 \\ \mathbf{c}_2 \end{pmatrix} \\ &= \begin{pmatrix} R^T & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \mathbf{c}_2 \end{pmatrix} = 0. \end{aligned}$$

Thus $\mathbf{r}_{LS}$ is orthogonal to the columns of X.

There are three well known ways to construct a Q–R decomposition:

- Householder–Golub factorization.
- Modified Gram–Schmidt orthogonalization
- Givens Q–R factorization.

We will say a lot about the first two and just touch on the third. MATLAB uses the first and we will start with that.

The matrix

$$H = I - 2\mathbf{w}\mathbf{w}^T, \quad \|\mathbf{w}\|_2 = 1$$

is called a *Householder* transformation. Often it is defined in terms of any nonzero vector $\mathbf{v}$ as

$$H = I - 2\mathbf{v}\mathbf{v}^T/(\mathbf{v}^T\mathbf{v}).$$

As you showed in your homework,

$$H = H^T, \quad H^T H = I, \quad H^2 = I.$$

It is common to choose H such that for a given vector $\mathbf{x}$,

$$H\mathbf{x} = \alpha\mathbf{e}_1 = \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Since H is orthogonal,

$$\|H\mathbf{x}\|_2 = \|\mathbf{x}\|_2 = |\alpha|\|\mathbf{e}_1\|_2 = |\alpha|.$$

We use this transformation to insert zeros into a matrix. The choice of $\mathbf{w}$ is

$$\mathbf{w} = (\mathbf{x} - \alpha\mathbf{e}_1)/\|\mathbf{x} - \alpha\mathbf{e}_1\|_2.$$

We note that

$$\begin{aligned}\|\mathbf{x} - \alpha\mathbf{e}_1\|_2^2 &= \mathbf{x}^T\mathbf{x} + \alpha^2\mathbf{e}_1^T\mathbf{e}_1 - 2\alpha\mathbf{x}^T\mathbf{e}_1 \\ &= \mathbf{x}^T\mathbf{x} + \alpha^2\mathbf{e}_1^T\mathbf{e}_1 - 2\alpha x_1\end{aligned}$$

To prevent cancellation in the first entry of the Householder vector, it is recommended to choose

$$\alpha = -\text{sign}(x_1)\|\mathbf{x}\|_2,$$

so that

$$\|\mathbf{x} - \alpha\mathbf{e}_1\|_2^2 = 2\|\mathbf{x}\|_2^2 + 2\|\mathbf{x}\|_2|x_1|.$$

If you go to my notes for CSE/Math 550, it shows a way to allow you to choose α to have the opposite sign `http://www.cse.psu.edu/~barlow/cse550/chap4.pdf`.

Show for yourself that

$$H\mathbf{x} = (I - 2\mathbf{w}\mathbf{w}^T)\mathbf{x} = \alpha\mathbf{e}_1.$$

To apply a Householder transformation, that is to compute

$$C = HB$$

we use

$$C = (I - 2\mathbf{w}\mathbf{w}^T)B = B - 2\mathbf{w}\mathbf{w}^T B = B - \mathbf{w}\mathbf{f}^T$$

where

$$\mathbf{f} = 2B^T\mathbf{w}.$$

Thus a Householder transformation is the result of a matrix–vector product and an outer product. The latter computes the components of C from

$$c_{ij} = b_{ij} - w_i f_j.$$

Although this is $4mn$ arithmetic operations, thus much less work than multiplying two matrices.

To compute the Q–R factorization of X, let

$$X = (\mathbf{x}_1, \ldots, \mathbf{x}_n).$$

Choose H_1, a Householder transformation such that

$$H_1\mathbf{x}_1 = r_{11}\mathbf{e}_1,$$

thus

$$X_1 = H_1 X = \begin{array}{c} \\ 1 \\ m-1 \end{array} \begin{array}{c} \begin{array}{cc} 1 & n-1 \end{array} \\ \begin{pmatrix} r_{11} & R_{12} \\ 0 & \tilde{X}_1 \end{pmatrix} \end{array}$$

where

$$\tilde{X}_1 = (\mathbf{x^{(1)}}_2, \ldots, \mathbf{x^{(1)}}_n).$$

Choose $\tilde{H}_2$ such that

$$\tilde{H}_2 \mathbf{x}_2^{(1)} = r_{22} \mathbf{e}_1.$$

Then let

$$H_2 = \begin{array}{c} \\ 1 \\ m-1 \end{array} \begin{array}{c} \begin{array}{cc} 1 & m-1 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & \tilde{H}_2 \end{pmatrix} \end{array}$$

which leaves the first row unaffected. Then

$$X_2 = H_2 X_1 = H_2 H_1 X = \begin{array}{c} \\ 2 \\ m-2 \end{array} \begin{array}{c} \begin{array}{cc} 2 & n-2 \end{array} \\ \begin{pmatrix} R_{11}^{(2)} & R_{12}^{(2)} \\ 0 & \tilde{X}_2 \end{pmatrix} \end{array}.$$

Suppose

$$\begin{aligned} X_{k-1} &= H_{k-1} \cdots H_1 X \\ &= \begin{array}{c} \\ k-1 \\ m-k+1 \end{array} \begin{array}{c} \begin{array}{cc} k-1 & n-k+1 \end{array} \\ \begin{pmatrix} R_{11}^{(k-1)} & R_{12}^{(k-1)} \\ 0 & \tilde{X}_{k-1} \end{pmatrix} \end{array} \end{aligned}$$

where

$$\tilde{X}_{k-1} = (\mathbf{x^{(k-1)}}_k, \ldots, \mathbf{x^{(k-1)}}_n).$$

Choose

$$\tilde{H}_k \mathbf{x}_k^{(k-1)} = r_{kk} \mathbf{e}_1$$

and let

$$H_k = \begin{pmatrix} I_{k-1} & 0 \\ 0 & \tilde{H}_k \end{pmatrix}.$$

Then

$$\begin{aligned} X_k &= H_k X_{k-1} = H_k \cdots H_1 X \\ &= \begin{array}{c} \\ k \\ m-k \end{array} \begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \begin{pmatrix} R_{11}^{(k-1)} & R_{12}^{(k-1)} \\ 0 & \tilde{X}_{k-1} \end{pmatrix} \end{array} \end{aligned}$$

At $k = n$ we have

$$X_n = H_n \cdots H_1 X = \begin{matrix} n \\ m-n \end{matrix} \begin{pmatrix} R \\ 0 \end{pmatrix}$$

where R is upper triangular and nonsingular. Thus

$$Q^T = H_n \cdots H_1$$

and

$$Q = H_1 \cdots H_n. \tag{2}$$

Also, MATLAB explicitly computes Q, that is not necessary. Instead we can store the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_n$ that define

$$H_k = I - 2\mathbf{w}_k \mathbf{w}_k^T.$$

Since the first $k-1$ components of $\mathbf{w}_k$ are zero, in many codes, X is overwritten by (in a 4×3 case)

$$\begin{matrix} w_{11} & r_{12} & r_{13} \\ w_{21} & w_{22} & r_{23} \\ w_{31} & w_{32} & w_{33} \\ w_{41} & w_{42} & w_{43} \end{matrix}.$$

The diagaonal of R can be stored in an extra vector $(r_{11}, r_{22}, r_{33})^T$, say.

Now, we can fill in some blanks. If Q is as in (2), we compute $\mathbf{c}$ from

$$\mathbf{c} = H_n \cdots H_1 \mathbf{b} = Q^T \mathbf{b}$$

and $\mathbf{r}_{LS}$ from

$$\mathbf{r}_{LS} = H_1 \cdots H_n \begin{pmatrix} 0 \\ \mathbf{c}_2 \end{pmatrix}.$$

Then we recover the solution from

$$R\mathbf{y}_{LS} = \mathbf{c}_1.$$

Range and Null Spaces The *range* of a matrix $X \in \mathbf{R}^{m \times n}$ is the set

$$\text{Range}(X) = \{X\mathbf{y} : \mathbf{y} \in \mathbf{R}^n\}.$$

The range is a subspace of $\mathbf{R}^m$. The *rank* of X is the dimension of its range space and is written rank(X). Clearly, rank$(X) \leq min\{m, n\}$. The *span* of a set of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is given by

$$\text{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} = \text{Range}(X)$$

where $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. If rank$(X) = n$, then $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a *basis* for Range(X) and X is a *basis matrix.* A matrix X is *rank deficient* if rank$(X) < min\{m, n\}$, is said to have *full column rank* if rank$(X) = n$, and is said to have *full row rank* if rank$(X) = m$.

The *null space* of X is the linear subspace of $\mathbf{R}^n$ given by

$$\text{Null}(X) = \{\mathbf{y} \in \mathbf{R}^n : X\mathbf{y} = 0\}.$$

The columns of Q contains bases for two important subspaces. Let

$$Q = \begin{matrix} & n & m-n \\ & \big(Q_1 & Q_2 \big) \end{matrix}.$$

The matrices $Q_1 \in \mathbf{R}^{m \times n}$ and $Q_2 \in \mathbf{R}^{m \times m-n}$ are left orthogonal matrices satisfying $Q_1^T Q_2 = 0$.

Since

$$X = Q_1 R$$

it is easily verified that

$$\text{Range}(Q_1) = \text{Range}(X).$$

That is, the columns of Q_1 are an orthonormal basis for Range(X). Since

$$X^T Q_2 = R^T Q_1^T Q_2 = 0$$

then one can show that

$$\text{Range}(Q_2) = \text{Range}(X)^\perp = (X^T).$$

From these two matrices we get orthogonal projections.

$$P_1 = Q_1 Q_1^T, \quad P_2 = Q_2 Q_2^T$$

are projections on the spaces Range(X) and Null(X^T).