MA 511, Session 17

Orthogonal Matrices

Definition: Suppose $v_1, \dots, v_k \in \mathbb{R}^n$. We say the vectors are orthonormal if $v_i^T v_j = \delta_{ij}$, the Kronecker symbol.

This means that they are pairwise *orthogonal*, and their norms are equal to 1 (*normal*).

Example: The standard basis of $\mathbb{R}^n$, $e_1, \dots, e_n$ is orthonormal.

Example:

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix}, v_2 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

are orthonormal.

Definition: A $n \times n$ matrix Q is <u>orthogonal</u> if the columns of Q are orthonormal vectors in $\mathbb{R}^n$.

Theorem: Let Q be orthogonal. Then,

(i) $Q^T = Q^{-1}$

(ii) the rows of Q are orthonormal

(iii) Q preserves lengths, i.e. $\|Qx\| = \|x\|$ for all $x \in \mathbb{R}^n$

(iv) Q preserves inner products, i.e. $(Qx)^T Qy = x^T y$ for all $x, y \in \mathbb{R}^n$

<u>Proof</u>: (i) follows from the definition: $Q^T Q = \mathbb{I}$

(ii) Follows from (i): $\mathbb{I} = QQ^{-1} = QQ^T$

(iii) $\|Qx\|^2 = (Qx)^T Qx = x^T Q^T Qx = x^T \mathbb{I} x = \|x\|^2$

(iv) $(Qx)^T Qy = x^T Q^T Qy = x^T \mathbb{I} y = x^T y$

Example: The rotation matrix

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

is orthogonal ($\sin^2\alpha + \cos^2\alpha = 1$).

Suppose now that a $m \times n$ matrix Q has orthonormal columns, with $m \geq n$, and let $b \in \mathcal{C}(Q)$. Then, the unique solution of

$$Qx = b$$

is given by

$$x = Q^T b,$$

since $x = \mathbb{I}x = Q^T Q x = Q^T b$, as we are assuming $Q^T Q = \mathbb{I}_n$ even though the matrix Q is not orthogonal unless $m = n$.

Remark: Assume the problem $Qx = b$ is inconsistent ($m > n$). Then, the least squares solution is also $\bar{x} = Q^T b$. Therefore,

(i) $p = Q\bar{x} = QQ^T b$ is the orthogonal projection of b onto the column space of Q, and $P = QQ^T$ is the projection matrix

(ii) $p = p_1 + \cdots + p_n$ where $p_j = (q_j^T b) q_j$ are projections of b onto the lines spanned by the columns q_j of Q

Example: Straight Line Approximation.

Let us try to fit the data

$$b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \quad \text{at times} \quad \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$$

to the linear model $C + Dt = b$. Then,

$$\begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

Note that the coefficient matrix

$$A = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{pmatrix}$$

has orthogonal columns (though not orthonormal) if $t_1 + t_2 + t_3 = 0$.

Let

$$T = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$$

satisfy this relation. Then, the least squares solution is given by

$$A^T A \begin{pmatrix} \bar{C} \\ \bar{D} \end{pmatrix} = A^T b,$$

where

$$A^T A = \begin{pmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \end{pmatrix} \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & t_1^2 + t_2^2 + t_3^2 \end{pmatrix}$$

and

$$\begin{pmatrix} 3 & 0 \\ 0 & t_1^2 + t_2^2 + t_3^2 \end{pmatrix} \begin{pmatrix} \bar{C} \\ \bar{D} \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 1 & 1 \\ t_1 & t_2 & t_3 \end{pmatrix} b = \begin{pmatrix} b_1 + b_2 + b_3 \\ b_1 t_1 + b_2 t_2 + b_3 t_3 \end{pmatrix},$$

finally giving

$$\bar{C} = \frac{(\,1 \quad 1 \quad 1\,)\, b}{3}, \qquad \bar{D} = \frac{T^T b}{\|T\|^2}.$$

In the general case,

$$\begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \dots & \dots \\ 1 & t_m \end{pmatrix} \begin{pmatrix} \bar{C} \\ \bar{D} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix} = b,$$

has the least squares solution

$$T = \begin{pmatrix} t_1 \\ t_2 \\ \dots \\ t_m \end{pmatrix}, \qquad \bar{C} = \frac{b_1 + \cdots + b_m}{m}, \quad \bar{D} = \frac{T^T b}{\|T\|^2},$$

if $t_1 + t_2 + \cdots + t_m = 0$.

Since these formulas are computationally so simple, it is always advantageous to make a preliminary change of variable $\hat{t} = t - \bar{t}$, where $\bar{t} = \frac{t_1 + \cdots + t_m}{m}$ is the average of the values $t_1, \dots, t_m$. In this way $\hat{t}_1 + \hat{t}_2 + \cdots + \hat{t}_m = 0$. Then the model becomes

$$b = C + D\hat{t},$$

that is,

$$b = C + D(t - \bar{t}).$$

Gram-Schmidt Orthonormalization Method

Since orthonormal vectors are easier to work with, we need an algorithm to change vectors $a_1, \ldots, a_k$ by linear combinations to an orthonormal set. The Gram-Schmidt method produces orthonormal vectors $q_1, \ldots, q_k$ such that the span of $q_1, \ldots, q_j$ is the same as the span of $a_1, \ldots, a_j$ for $1 \le j \le k$.

We do this by iteratively subtracting off projections onto previous subspaces:

$$a'_j = a_j - p_{j-1} = a_j - (q_1^T a_j)q_1 - \cdots - (q_{j-1}^T a_j)q_{j-1},$$

where p_{j-1} is the projection of a_j onto the subspace spanned by $q_1, \ldots, q_{j-1}$. Subsequently a'_j is discarded if it is 0, or normalized if it is not:

$$q_j = \frac{a'_j}{\|a'_j\|}.$$

If $a_1, \ldots, a_k$ are linearly independent, we do not need to worry about discarding any vectors, they will be linearly independent and thus, *a fortiori*, nonzero.

Example: Apply the Gram-Schmidt method to

$$a_1 = \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}, \ a_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \ a_3 = \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}.$$

<u>Solution</u>: $q_1 = \frac{1}{5} \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}$. Next,

$$\begin{aligned} a_2' &= \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} - \begin{pmatrix} \frac{3}{5} & \frac{4}{5} & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \begin{pmatrix} \frac{3}{5} \\ \frac{4}{5} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} - \frac{4}{5} \begin{pmatrix} \frac{3}{5} \\ \frac{4}{5} \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{12}{25} \\ \frac{9}{25} \\ 2 \end{pmatrix} \quad \text{gives} \\ q_2 &= \frac{1}{\sqrt{109}} \begin{pmatrix} -\frac{12}{5} \\ \frac{9}{5} \\ 10 \end{pmatrix}. \end{aligned}$$

Finally,

$$
\begin{aligned}
a_3' &= \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix} - \begin{pmatrix} \frac{3}{5} & \frac{4}{5} & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix} \begin{pmatrix} \frac{3}{5} \\ \frac{4}{5} \\ 0 \end{pmatrix} \\
&\quad - \frac{1}{109} \begin{pmatrix} -\frac{12}{5} & \frac{9}{5} & 10 \end{pmatrix} \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix} \begin{pmatrix} -\frac{12}{5} \\ \frac{9}{5} \\ 10 \end{pmatrix} \\
&= \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix} - \frac{29}{25} \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix} - \frac{1}{5} \begin{pmatrix} -\frac{12}{5} \\ \frac{9}{5} \\ 10 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},
\end{aligned}
$$

which was to be expected since a_3 is a linear combination of a_1 and a_2 ($a_1 + a_2 = a_3$); this means that the projection of a_3 onto the subspace spanned by a_1 and a_2 is a_3 itself, and this is what we subtract from a_3 to create a_3'. Therefore, the Gram-Schmidt method produces only 2 vectors, and it eliminates the third.

The QR Decomposition of a Matrix.

Let A be a $m \times n$ matrix with $m \geq n$ and assume rank $(A) = n$. We shall produce a factorization $A = QR$, where Q is a $m \times n$ matrix with orthonormal columns, and R is an upper triangular $n \times n$ matrix.

Since the n columns of A, $C_1, \ldots, C_n$, are linearly independent vectors in $\mathbb{R}^m$, we can apply the Gram-Schmidt method to them and produce n orthonormal vectors $q_1, \ldots, q_n \in \mathbb{R}^m$ that will be the columns of the matrix Q. The coefficients of R are the projections of the columns of A onto the lines of direction $q_1, \ldots, q_n$. From the Gram-Schmidt method we have (for $n = 3$)

$$\begin{pmatrix} C_1 & C_2 & C_3 \end{pmatrix} = \begin{pmatrix} q_1 & q_2 & q_3 \end{pmatrix} \begin{pmatrix} q_1^T C_1 & q_1^T C_2 & q_1^T C_3 \\ 0 & q_2^T C_2 & q_2^T C_3 \\ 0 & 0 & q_3^T C_3 \end{pmatrix},$$

since

$$\begin{aligned} C_1 &= q_1^T C_1 q_1 \\ C_2 &= q_1^T C_2 q_1 + q_2^T C_2 q_2 \\ C_3 &= q_1^T C_3 q_1 + q_2^T C_3 q_2 + q_3^T C_3 q_3. \end{aligned}$$