Orthogonal Bases and Gram-Schmidt(标准正交基和施密特正交化过程)

Lecture 16

Dept. of Math., SUSTech

2023.11

Orthogonal Bases and Gram-Schmidt

Orthogonal Vectors

- In an orthogonal basis, every vector is perpendicular to every other vector.
- The coordinate axes are mutually orthogonal. That is just about optimal, and the one possible improvement is easy: Divide each vector by its length, to make it a unit vector.

Definition

The vectors $q_1, q_2, \cdots, q_n$ are orthonormal if

$$q_i^T q_j = \begin{cases} 0 & \text{whenever } i \neq j, \text{giving the orthogonality;} \\ 1 & \text{whenever } i = j, \text{giving the normalization.} \end{cases}$$

A matrix with orthonormal columns will be called Q.

Introduction

If we have a subspace of $\mathbb{R}^n$, the standard vectors e_i might not lie in that subspace. But the subspace always has an orthonormal basis, and it can be constructed in a simple way out of any basis whatsoever. This construction, which **converts a skewed set of axes into a perpendicular set**, is known as **Gram-Schmidt orthogonalization**. To summarize, the three topics basic to this section are:

1. The definition and properties of orthogonal matrices Q.
2. The solution of $Qx = b$, either n by n or rectangular (least-squares).
3. The Gram-Schmidt process and its interpretation as a new factorization $A = QR$.

Orthogonal Matrices

Proposition

If Q (square or rectangular) has orthonormal columns, then $Q^TQ = I$:

$$\begin{bmatrix} q_1^T \\ q_2^T \\ \dots \\ q_n^T \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} = I$$

An orthogonal matrix is a square matrix with orthonormal columns. Then $Q^T = Q^{-1}$. For square orthogonal matrices, the transpose is the inverse.

Note that $Q^TQ = I$ even if Q is rectangular. But then Q^T is only a left-inverse.

Examples

Example

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad Q^T = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Q rotates every vector through the angle θ, and Q^T rotates it back through $-\theta$. The columns are clearly orthogonal, and they are orthonormal because $\sin^2\theta + \cos^2\theta = 1$. The matrix Q^T is just as much an orthogonal matrix as Q.

Example

Any permutation matrix P is an orthogonal matrix. Geometrically, an orthogonal Q is the product of a rotation and a reflection.

Length Preserving

There does remain one property that is shared by rotations and reflections, and in fact by every orthogonal matrix.

Proposition

Multiplication by any Q preserves lengths:

$$||Qx|| = ||x|| \quad \textit{for every vector} \quad x$$

Remarks:

- This property is not shared by projections, which are not orthogonal or even invertible.
- Projections reduce the length of a vector, whereas orthogonal matrices preserve lengths.

Linear Combinations

Write b as a combination

$$b = x_1 q_1 + x_2 q_2 + \cdots + x_n q_n,$$

to find x_i, we multiply both sides of the equation by q_i^T.

Remarks:

- Every vector b is the sum of its one-dimensional projections onto the lines through the q's.
- The rows of a square matrix are orthonormal whenever the columns are. Example:

$$Q = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

Rectangular matrices with Orthogonal Columns

If Q has orthonormal columns, the least-squares problem becomes easy: rectangular system with no solution for most b and the projection matrix is $P = QQ^T$.

$Qx = b$ rectangular system with no solution for most b

$Q^T Q\hat{x} = Q^T b$ normal equations for the best $\hat{x}$ $-$ in which $Q^T Q = I$.

$\hat{x} = Q^T b$ $\hat{x}_i$ is $q_i^T b$.

$p = Q\hat{x}$ the projection of b is $(q_1^T b)q_1 + \cdots + (q_n^T b)q_n$.

$p = QQ^T b$ the projection matrix is b is $P = QQ^T$.

Remarks

The last formulas are like $p = A\hat{x}$ and $P = A(AA^T)^{-1}A^T$. When the columns are orthonormal, the “cross product matrix” A^TA becomes $Q^TQ = I$. The projections onto the axes are uncoupled, and p is the sum

$$p = (q_1^T b)q_1 + \cdots + (q_n^T b)q_n.$$

Examples

Example

Project a point $b = (x, y, z)$ onto the x-y plane.

Remark:

Projection onto a plane=sum of projections onto orthonormal q_1 and q_2.

Example

Example

Example 4 When the measurement times average to zero, fitting a straight line leads to orthogonal columns. Take $t_1 = -3, t_2 = 0$, and $t_3 = 3$. Then the attempt to fit $y = C + Dt$ leads to three equations in two unknowns:

$$\left\{ \begin{array}{l} C + Dt_1 = y_1 \\ C + Dt_2 = y_2 \\ C + Dt_3 = y_3 \end{array} \right. \quad \text{or} \quad \begin{bmatrix} 1 & -3 \\ 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

The columns $(1,1,1)$ and $(-3,0,3)$ are orthogonal. We can project y separately onto each column, and the best coefficients $\widehat{C}$ and $\widehat{D}$ can be found separately.

The Gram-Schmidt Process

- Suppose you are given three independent vectors a, b, c. If they are orthonormal, life is easy. If they are not orthonormal, we need to propose a way to make them orthonormal.
- The idea is to subtract from every new vector its components in the directions that are already settled.

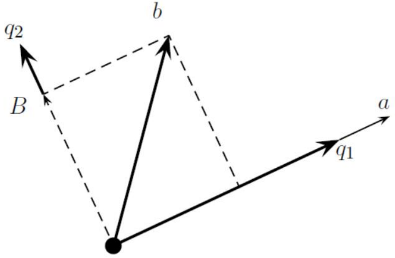

Figure 3.10: The q_i component of b is removed; a and B normalized to q_1 and q_2.

The Gram-Schmidt Process

- The Gram-Schmidt process starts with independent vectors $a_1, a_2, \cdots, a_n$ and ends with orthonormal vectors $q_1, q_2, \cdots, q_n$.
- At step j it subtracts from a_j its components in the directions $q_1, \cdots, q_{j-1}$ that are already settled:

$$A_j = a_j - (q_1^T a_j) q_1 - \cdots - (q_{j-1}^T a_j) q_{j-1}.$$

- Then q_j is the unit vector

$$q_j = \frac{A_j}{||A_j||}.$$

Example: Gram-Schmidt

Example 5. Gram-Schmidt Suppose the independent vectors are a, b, c:

$$a = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad c = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

$$B = b - (q_1^{\mathrm{T}} b) q_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

$$\begin{aligned} C &= c - (q_1^{\mathrm{T}} c) q_1 - (q_2^{\mathrm{T}} c) q_2 \\ &= \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \sqrt{2} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} - \sqrt{2} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \end{aligned}$$

Orthonormal basis $\quad Q = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix}.$

The Factorization $A = QR$

- Every m by n matrix with independent columns can be factored into $A = QR$.
- The columns of Q are orthonormal, and R is upper triangular and invertible.
- When $m = n$ and all matrices are square, Q becomes an orthogonal matrix.

Remark on the calculations:

- It is easier to compute the orthogonal a, B, C, without forcing their lengths to equal one.

QR:Example

We started with a matrix A, whose columns were a, b, c. We ended with a matrix Q, whose columns are q_1, q_2, q_3. The QR factorization is as follows:

$$QR = \begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} q_1^T a & q_1^T b & q_1^T c \\ 0 & q_2^T b & q_2^T c \\ 0 & 0 & q_3^T c \end{bmatrix}$$

From example 5, we deduce that:

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix} = QR.$$

- You see the lengths of a, B, C on the diagonal of R.
- The orthonormal vectors q_1, q_2, q_3, which are the whole object of orthogonalization, are in the first factor Q.

QR factorization

Maybe QR is not as beautiful to the theory as LU (because of the square roots). Both factorizations are vitally important to the theory of linear algebra, and absolutely central to the calculations. If LU is Hertz, then QR is Avis.

Theorem

Every m by n matrix with independent columns can be factored into $A = QR$. The columns of Q are orthonormal, and R is upper triangular and invertible. When $m = n$ and all matrices are square, Q becomes an orthogonal matrix.

Function Spaces and Fourier Series

1. Hilbert Space.
2. Lengths and Inner Products.
3. Fourier Series.
4. Gram-Schmidt for Functions.
5. Best Straight Line.

Homework Assignment 16

3.4: 1, 2, 5, 6, 11, 13, 17, 28, 30.