MATH-UA 140- Linear Algebra

Lecture 17: Orthogonal Bases, Orthogonal Matrices, and QR decomposition

In the previous lecture, we saw that computing the least square approximation $\hat{\vec{x}}$ was particularly easy when the measurement times added to 0 because A^TA was diagonal.
A central purpose of the present lecture is to learn how to express a matrix A for which A^TA is not diagonal in terms of a new matrix Q for which Q^TQ is diagonal, thereby rendering the calculation of $\hat{\vec{x}}$ more straightforward.

I) Orthonormal bases and orthogonal matrices

We already know that n basis vectors $\vec{q}_1, \dots, \vec{q}_n$ are called orthogonal if $\vec{q}_i^{\,T}\vec{q}_j = 0$ for $i \neq j$. If all the basis vectors are unit vectors, the basis $\vec{q}_1, \dots, \vec{q}_n$ is called orthonormal:

1) Definition

The vectors $\vec{q}_1, \dots, \vec{q}_n$ are orthonormal if

$$\vec{q}_i^{\,T}\vec{q}_j = \begin{cases} 0 \text{ when } i \neq j & \text{(orthogonality condition)} \\ 1 \text{ when } i = j & \text{(unit length for each } \vec{q}_i\text{)} \end{cases}$$

2) Matrices with orthonormal columns

Consider n orthonormal vectors $\vec{q}_1, \dots, \vec{q}_n$, and the matrix $Q = [\vec{q}_1 \cdots \vec{q}_n]$ in which the $\vec{q}_i$ are the columns of Q.

Let us evaluate $Q^T Q = \begin{bmatrix} \vec{q}_1^{\,T} \\ \vdots \\ \vec{q}_n^{\,T} \end{bmatrix} [\vec{q}_1 \cdots \vec{q}_n]$

$$= \begin{bmatrix} \vec{q}_1^{\,T}\vec{q}_1 & \vec{q}_1^{\,T}\vec{q}_2 & \cdots & \vec{q}_1^{\,T}\vec{q}_n \\ \vdots & \vdots & & \vdots \\ \vec{q}_n^{\,T}\vec{q}_1 & \vec{q}_n^{\,T}\vec{q}_2 & \cdots & \vec{q}_n^{\,T}\vec{q}_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} = I$$

A matrix Q with orthonormal columns satisfies $Q^T Q = I$

3) Orthogonal matrices

When Q is a square matrix, $Q^T Q = I$ implies that $Q^{-1} = Q^T$

Square matrices with orthonormal columns are called orthogonal matrices. Orthogonal matrices satisfy $Q^T = Q^{-1}$

Note: We all agree that it would have made much more sense to call any matrix with orthonormal columns (whether the matrix is square or not) an "orthonormal matrix". For some reason, noone is using this terminology. Also, orthogonal matrix is reserved for square matrices.

4) Matrices with orthonormal columns and dot product

Let Q be a matrix with n orthonormal columns, and $\vec{x}$ and $\vec{y}$ two vectors in $\mathbb{R}^n$. Q preserves the dot product of $\vec{x}$ and $\vec{y}$, in the sense that:

$$(Q\vec{x})\cdot(Q\vec{y}) = (Q\vec{x})^T(Q\vec{y}) = \vec{x}^T \underbrace{Q^TQ}_{I}\vec{y} = \vec{x}^T\vec{y} = \vec{x}\cdot\vec{y}$$

In particular, Q preserves lengths:

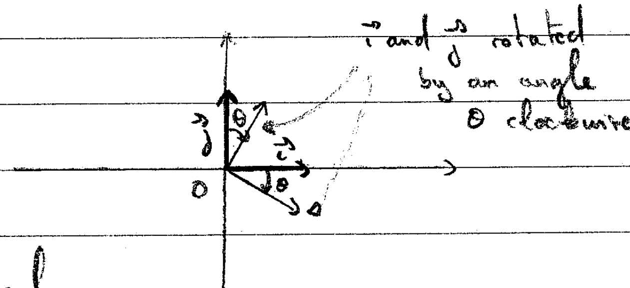
$$\|Q\vec{x}\|^2 = (Q\vec{x})^T Q\vec{x} = \vec{x}^T \underbrace{Q^T Q}_{I}\vec{x} = \vec{x}^T\vec{x} = \|\vec{x}\|$$

We will now show examples to give a sense of what matrices with orthonormal columns may be, and you will more intuitively visualize the dot product/length preservation property.

5) Examples

A) Consider the two-dimensional x-y plane, and the matrix Q which rotates vectors about the origin by an angle θ in the clockwise direction. Q can be written as:

$$Q = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$



The columns of Q are orthogonal.
Their length is 1. They form an orthonormal basis of $\mathbb{R}^2$, and Q is an orthogonal matrix, $Q^TQ = I$.
The inverse of Q is the rotation with angle $-\theta$. This provides

a simple way to see that $Q^T = Q^{-1}$.

B) Consider the permutation P such that

$$\underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}}_{P} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} z \\ y \\ x \end{bmatrix}$$

The columns of P are $\vec{k}, \vec{j}, \vec{i}$, clearly an orthonormal basis of $\mathbb{R}^3$. So P is an orthogonal matrix. Its inverse is P^T.

$P^T = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, which puts $\begin{bmatrix} z \\ y \\ x \end{bmatrix}$ back in the right order. In this case, we even have $P = P^{-1}$. This is however not true for any permutation matrix. What is true is that <u>every permutation matrix is an orthogonal matrix</u>, since a permutation matrix has the columns of the identity in a different order.

II] Constructing an orthonormal basis and QR decomposition

1) The Gram-Schmidt Process

Imagine one is given three linearly independent vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$. We will now learn how to create three orthonormal vectors $\vec{q}_1$, $\vec{q}_2$, and $\vec{q}_3$ out of $\vec{a}, \vec{b}, \vec{c}$. The key step is to find three orthogonal vectors $\vec{A}$, $\vec{B}$, and $\vec{C}$ from $\vec{a}, \vec{b}$, and $\vec{c}$.

From $\vec{A}$, $\vec{B}$, and $\vec{C}$ one simply construct $\vec{q_1}$, $\vec{q_2}$, and $\vec{q_3}$ by dividing by their respective magnitudes.

The construction of $\vec{A}$, $\vec{B}$, and $\vec{C}$ is known as the Gram-Schmidt process, and works as follows:

A) Pick $\vec{a} = \vec{A}$ - this amounts to choosing a first direction arbitrarily

B) Take $\vec{b}$ and decompose it into its projection onto $\vec{A}$ and a vector perpendicular to it. The latter is $\vec{B}$:

$$\vec{B} = \vec{b} - \frac{\vec{A}^T \vec{b}}{\vec{A}^T \vec{A}} \vec{A}$$

By construction, $\vec{A}$ and $\vec{B}$ are orthogonal. You can also verify this by direct computation.

C) Take $\vec{c}$ and decompose it into its projection onto $\vec{A}$, its projection onto $\vec{B}$, and a vector orthonormal to both. The latter is $\vec{C}$:

$$\vec{C} = \vec{c} - \frac{\vec{A}^T \vec{c}}{\vec{A}^T \vec{A}} \vec{A} - \frac{\vec{B}^T \vec{c}}{\vec{B}^T \vec{B}} \vec{B}$$

That's it, we have $\vec{A}$, $\vec{B}$, and $\vec{C}$ orthogonal to one another. The idea is always the same. If we had 4 vectors $\vec{a}$, $\vec{b}$, $\vec{c}$, and $\vec{d}$, we would subtract the three projections of $\vec{d}$ onto $\vec{A}$, $\vec{B}$, and $\vec{C}$ to obtain $\vec{D}$.

The final step is to divide $\vec{A}$, $\vec{B}$, and $\vec{C}$ by their respective magnitudes to get the <u>orthonormal</u> vectors $\vec{q_1}$, $\vec{q_2}$, and $\vec{q_3}$.

Example

$$\vec{a} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{c} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{A} = \vec{a} \Rightarrow \vec{A}^T \vec{A} = 2$$

$$\vec{B} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix}$$

$$\vec{B}^T \vec{B} = \frac{3}{2}$$

$$\vec{C} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{2}{3} \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix}$$

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

Hence, $\vec{q}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{q}_2 = \sqrt{\frac{2}{3}} \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$

$$\vec{q}_3 = \frac{\sqrt{3}}{2} \begin{bmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

2) Matrix form of the Gram-Schmidt Process: QR decomposition

We started with a matrix A, whose columns are the vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$, and the Gram-Schmidt process gave us a matrix Q whose columns are the orthonormal vectors $\vec{q_1}, \vec{q_2}, \vec{q_3}$. Since the only operations involved in the process were addition/subtraction, dot products, and divisions by scalars, there is a matrix R such that

$A = QR$

Now, remember that any column j of the matrix QR can be viewed as the linear combination $r_{1j}\vec{q_1} + r_{2j}\vec{q_2} + r_{3j}\vec{q_3}$, where the r_{ij}'s are the entries of the matrix R.

The first column of A, $\vec{a}$, is aligned with $\vec{q_1}$ and does not depend on $\vec{q_2}$ and $\vec{q_3}$. Hence, $r_{21} = r_{31} = 0$
The second column of A, $\vec{b}$, is in the plane of $\vec{q_1}$ and $\vec{q_2}$, and does not depend on $\vec{q_3}$: $r_{32} = 0$.

We see that by the nature of the Gram-Schmidt process, R must be upper triangular.

This is of course true no matter the number of columns A has.

Now, since Q has orthonormal columns, $Q^T Q = I$. Multiplying $A = QR$ by Q^T on the left on both sides, we obtain

$Q^T A = R$

which gives the explicit representation of the QR decomposition:

$$\begin{bmatrix} & & \\ \vec{a} & \vec{b} & \vec{c} \\ & & \end{bmatrix} = \begin{bmatrix} & & \\ \vec{q}_1 & \vec{q}_2 & \vec{q}_3 \\ & & \end{bmatrix} \begin{bmatrix} \vec{q}_1^{\,T}\vec{a} & \vec{q}_1^{\,T}\vec{b} & \vec{q}_1^{\,T}\vec{c} \\ 0 & \vec{q}_2^{\,T}\vec{b} & \vec{q}_2^{\,T}\vec{c} \\ 0 & 0 & \vec{q}_3^{\,T}\vec{c} \end{bmatrix}$$

$$\underline{A = QR}$$

and its straight forward generalization if A has more than 3 columns.

R is upper triangular because $\vec{a}$ is orthogonal to $\vec{q}_2$ and $\vec{q}_3$, and $\vec{b}$ to $\vec{q}_3$. The diagonal entries of R are the magnitudes of the associated $\vec{A}$, $\vec{B}$, and $\vec{C}$. They are always positive. Any $m \times n$ matrix with independent columns can be factored into QR.

Example:

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

III) Projections using orthonormal bases

1) Projection onto a subspace spanned by an orthonormal basis

Consider the matrix Q with orthonormal columns $\vec{q}_1, \ldots, \vec{q}_n$ spanning an n-dimensional subspace of $\mathbb{R}^m$

Imagine one is interested in the projection of a vector $\vec{b}$ of $\mathbb{R}^m$ onto this subspace.

We know the general formulas for the projection:

$$A^T A \hat{\vec{x}} = A^T \vec{b} \qquad \vec{p} = A(A^T A)^{-1} A^T \vec{b} \qquad P = A(A^T A)^{-1} A^T$$

Now, when $A = Q$, $A^T A = Q^T Q = I$, so these formulas simplify dramatically:

$$\hat{\vec{x}} = Q^T \vec{b} \qquad \vec{p} = Q Q^T \vec{b} \qquad P = Q Q^T$$

There are no matrices to invert! This is the main motivation for using orthonormal bases. The best $\hat{\vec{x}}$ is simply obtained by taking dot products of $\vec{q}_1, \ldots, \vec{q}_n$ with $\vec{b}$: the projection reduces to n 1-dimensional projections. Looking into the actual matrix entries, we recover the expansion we have always naturally used in other classes:

$$\vec{p} = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \end{bmatrix} \begin{bmatrix} \vec{q}_1^{\,T} \vec{b} \\ \vdots \\ \vec{q}_n^{\,T} \vec{b} \end{bmatrix} = (\vec{q}_1^{\,T} \vec{b}) \vec{q}_1 + (\vec{q}_2^{\,T} \vec{b}) \vec{q}_2 + \cdots + (\vec{q}_n^{\,T} \vec{b}) \vec{q}_n$$

For example, if $\vec{q_1}$, $\vec{q_2}$, and $\vec{q_3}$ are the basis $\vec{i}, \vec{j}, \vec{k}$ of $\mathbb{R}^3$, this is:

$$\vec{b} = (\vec{i}^T \vec{b})\vec{i} + (\vec{j}^T \vec{b})\vec{j} + (\vec{k}^T \vec{b})\vec{k} = (\vec{i}\cdot\vec{b})\vec{i} + (\vec{j}\cdot\vec{b})\vec{j} + (\vec{k}\cdot\vec{b})\vec{k}$$

our natural way of writing $\vec{b}$ in the basis $(\vec{i}, \vec{j}, \vec{k})$.

2) Projection and QR decomposition

Imagine we want to project a vector $\vec{b}$ onto a basis which is not orthonormal, to obtain the least square approximation $\hat{\vec{x}}$ for example. Let A be the matrix whose columns are the nonorthonormal basis. Then the QR decomposition of A simplifies the computation of $\hat{\vec{x}}$ significantly:

$$A^T A \hat{\vec{x}} = A^T \vec{b} \iff (QR)^T QR\hat{\vec{x}} = (QR)^T \vec{b}$$

$$\iff R^T \underbrace{Q^T Q}_{I} R \hat{\vec{x}} = R^T Q^T \vec{b}$$

$$\iff R^T R \hat{\vec{x}} = R^T Q^T \vec{b}$$

Now R^T is lower triangular with nonzero entries on the diagonal, so it is invertible. We can multiply the equality above by $(R^T)^{-1}$ on the left on both sides, to find:

$$R\hat{\vec{x}} = Q^T \vec{b}$$

Since R is upper triangular, this is quickly solved for $\hat{\vec{x}}$ by back substitution.