Rotating the Axes

Orthonormal Basis, Orthogonal Matrices, Gram-Schmidt and QR

Javier Rodríguez Laguna, UC3M

Last Update: March 6, 2008

Orthonormal Basis

In any vector space with an inner product, a set of vectors $\{u_i\}_{i=1}^n$ makes up an **orthogonal set** if they're all mutually perpendicular. In that case, $\langle u_i, u_j \rangle = 0$ if $i \neq j$. If each of them has unit norm, then the set is said to be **orthonormal**. In that case, we can say

$$\langle u_i, u_j \rangle = \delta_{ij}$$

The right hand side is the Kronecker's delta, which is one when $i = j$ and zero otherwise. Therefore, this equation is really saying that the inner product of two vectors of the set is zero when they're different, and one when they're the same.

If an orthogonal or orthonormal set makes up a basis for the vector space, we'll say it is an orthogonal or orthonormal basis. That's easy...

How to imagine an orthonormal basis? OK, all of its vectors are perpendicular and have unit norm. Therefore, they're just like the canonical basis, only rotated! (and, perhaps, reflected on some mirror). Why should we worry about them? Because they are much nicer than others! The reason is the following. Suppose we have a vector v and want to express it in a certain orthonormal basis $\{u_i\}_{i=1}^n$,

$$v = c_1 u_1 + \cdots + c_n u_n = \sum_{i=1}^n c_i u_i$$

How to obtain the c_i? It's very easy:

$$c_i = \langle v, u_i \rangle$$

Why? OK, just try to compute it!

$$\langle v, u_i \rangle = \left\langle \sum_{j=1}^n c_j u_j, u_i \right\rangle = \sum_{j=1}^n c_j \langle u_j, u_i \rangle$$

[Notice that we've changed the summation index to j, so as not to confuse it with i.]

But the fundamental feature of an orthonormal set is that all those inner products are zero but for the one corresponding to u_i, which is one. So we get the desired result. In symbols,

$$\langle v, u_i \rangle = \sum_{j=1}^n c_j \delta_{j,i} = c_i$$

Orthogonal Matrices

If you take the vectors u_i of an orthonormal basis and make up a matrix with them, putting them as columns, $U = [u_1, \cdots, u_n]$, then this matrix will implement the transformation in which e_1 (the first vector of the canonical basis) goes to u_1, e_2 goes to u_2 and so on. As we said before, this transformation is, in general terms, a rotation (possibly with reflections). These matrices are normally called **orthogonal matrices** (yes, yes, they should be called orthonormal, but what to do? this is the standard name!)

The most important property of orthogonal matrices come from the study of $U^T U$. Remember that the columns of U are the vectors u_i. Equivalently, the *rows* of U^T are the vectors u_i. So, now we want to get element ij of $U^T U$. We have to take column j of U, which is u_j. Then, we lay it horizontal and put it over row i of U^T, which is u_i. Now, we multiply and add, i.e.: we take the dot product! But that is 0 if $i \neq j$, and 1 if they're equal. So, we get the identity matrix!

$$U^T U = I$$

The proof can be reversed, and you get that a matrix is orthogonal if and only if its inverse coincides with its transpose. (Obviously, if the inner product is not the dot product, you have to take the adjoint instead of the transpose: $U^\dagger U = I$.)

E1. An example. Vectors $(1, 1)$ and $(1, -1)$ are orthogonal. Normalize them so that they make up an orthonormal basis of $\mathbf{R}^2$. Now, put them as columns of matrix U and check that $U^T U = I$.

Since the transformation $v \mapsto Uv$ is a rotation (with possible reflections), the length of v must be the same as that of Uv. We will state the more general property that $\langle Ux, Uy\rangle = \langle x, y\rangle$.

E2. (Theoretical) To get some practical with summations, try to prove $Ux \cdot Uy = x \cdot y$, in the case of $\mathbf{R}^n$ with the usual dot product. You'll have only to express it in terms of summations, and do a strategic interchange...

Let's try to prove that in general. Let's return to the idea of orthonormal basis. Then, if we say that

$$v = \sum_{i=1}^{n} c_i u_i$$

Then it is true that $v = Uc$, where c is the vector made from putting all the c_i in a column, right? Let's say then also that $w = Ud$, where a similar equation holds. Now, let's compute

$$\langle v, w\rangle = \langle Uc, Ud\rangle$$

Let's develop that last thing. If we reach $\langle c, d\rangle$, then our theorem is proved! Let's go.

$$\langle Uc, Ud\rangle = \left\langle \sum_i c_i u_i, \sum_j d_j u_j \right\rangle$$

Notice that we try not to repeat indices! Now, let's take out of the inner product the summations and the constants:

$$\langle Uc, Ud\rangle = \sum_{i,j} c_i d_j \langle u_i, u_j\rangle = \sum_{i,j} c_i d_j \delta_{i,j} = \sum_i c_i d_i$$

and that's the dot product of vectors c and d!

E3. Use the orthogonal matrix obtained in exercise 1 and check, for any two vectors x and y that $\langle x, y\rangle = \langle Ux, Uy\rangle$.

Some Important Orthogonal Matrices

A couple of types of orthogonal matrices are very important. The first one is **rotation matrices**. Suppose you're on $\mathbf{R}^2$. Then a counterclockwise rotation of angle α is implemented by

$$R(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

E4. Prove it with a picture.

E5. Dig up your trigonometry and check that $R(\alpha_1)R(\alpha_2) = R(\alpha_1 + \alpha_2)$.

If you're working in $\mathbf{R}^3$, then a rotation around the Z axis is written as

$$R_z(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

E6. What do you think should be the matrix for a rotation around the X or Y axis?

E7. Reflections on a coordinate plane in $\mathbf{R}^3$ can be achieved very easily... How?

Gram-Schmidt Orthonormalization

So, orthonormal basis are nice... What if we have a basis for a certain subspace but it is not orthonormal? In principle, all basis should be equivalent, no? So, perhaps, just for computational purposes, it would be nicer to change it a little bit, so that it becomes orthonormal. We can do it, the process is called **Gram-Schmidt orthonormalization**.

The Gram-Schmidt process starts with a set of vectors $\{a_i\}_{i=1}^m$ and produces a new set of vectors $\{u_i\}_{i=1}^m$ in such a way that the subspace they span is the same, but the new set is orthonormal. How to do it? Imagine that the set has a single vector. Then, it would be OK just to normalize it. So,

$$u_1 = \frac{a_1}{\|a_1\|}$$

Now, imagine that a second vector appears. We want to create a u_2 which is a linear combination of a_2 and a_1, but which is orthogonal to u_1. Wait, it can be also made with a_2 and u_1, it's equivalent, no? So, let's try the following:

$$\hat{u}_2 = a_2 - \langle a_2, u_1\rangle\, u_1$$

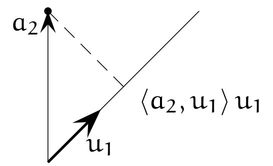

FIGURE 1. Obtaining $\hat{u}_2$.

Why the hat? Because that vector is not yet normalized! And why that expression? What is it? It's a projection! You substract from a_2 its contribution from u_1. See figure 1. The difference must be orthogonal to u_1!

Now, if you compute the inner product of $\hat{u}_2$ and u_1, it's coming like this:

$$\langle \hat{u}_2, u_1 \rangle = \langle a_2 - \langle a_2, u_1 \rangle u_1, u_1 \rangle = \langle a_2, u_1 \rangle - \langle a_2, u_1 \rangle \langle u_1, u_1 \rangle = 0$$

because $\langle u_1, u_1 \rangle = 0$! Now, we normalize $\hat{u}_2$ in a trivial way, and got the second vector: $u_2 = \hat{u}_2 / \|\hat{u}_2\|$.

E8. Orthonormalize the set of two vectors in $\mathbf{R}^3$, $a_1 = (1, 1, 0)$ and $a_2 = (0, 1, 1)$.

What to do with a third vector? We have to substract the contribution from u_1 and u_2, so:

$$\hat{u}_3 = a_3 - \langle a_3, u_2 \rangle u_2 - \langle a_3, u_1 \rangle u_1$$

Then, of course, you normalize that vector, $u_3 = \hat{u}_3 / \|\hat{u}_3\|$. You can continue with more vectors, always substracting the contribution from all the previous ones:

$$\hat{u}_k = a_k - \sum_{j=k-1}^{n} \langle a_k, u_j \rangle u_j$$

Let's do an example. In $\mathbf{R}^4$, let's orthonormalize the set $a_1 = (1, 1, 0, 0)$, $a_2 = (0, 1, 1, 0)$ and $a_3 = (0, 0, 1, 1)$. We start with u_1:

$$u_1 = \frac{1}{\sqrt{2}}(1, 1, 0, 0)$$

Now, we substract from a_2 its contribution on u_1:

$$\hat{u}_2 = a_2 - \langle a_2, u_1 \rangle u_1 = (0, 1, 1, 0) - \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}}(1, 1, 0, 0) = (-1/2, 1/2, 1, 0)$$

You can readily check that $\hat{u}_2$ is orthogonal to a_1. Now, let's normalize u_2:

$$u_2 = \frac{\hat{u}_2}{\|\hat{u}_2\|} = \sqrt{2/3}(-1/2, 1/2, 1, 0)$$

Last step: we compute $\hat{u}_3$:

$$\begin{aligned}\hat{u}_3 =& a_3 - \langle a_3, u_2 \rangle u_2 - \langle a_3, u_1 \rangle u_1 = \\ =&(0, 0, 1, 1) - \sqrt{2/3}\sqrt{2/3}(-1/2, 1/2, 1, 0) - 0\sqrt{1/2}(1, 1, 0, 0) \\ =&(1/3, -1/3, 1/3, 1)\end{aligned}$$

We normalize it and we have finished:

$$u_3 = \frac{\hat{u}_3}{\|\hat{u}_3\|} = \frac{\sqrt{3}}{2}(1/3, -1/3, 1/3, 1)$$

E9. Check that u_1, u_2, u_3 make up an orthonormal set.

E10. Apply the Gram-Schmidt process to orthonormalize the set in $\mathbf{R}^4$ given by $(1, 1, 1, 1)$, $(0, 1, 1, 1)$ and $(0, 0, 1, 1)$.

E11. Apply the Gram-Schmidt procedure to the set of vectors $\{1, x, x^2\}$ of the vector space $C[0, 1]$.

QR Factorization

The Gram-Schmidt process can give us a factorization of a matrix which is very useful. Any $m \times n$ matrix A with linearly independent columns can be written as QR, where Q is an $m \times n$ matrix whose columns are an orthonormal basis for Col(A) and R is an upper-triangular $n \times n$ invertible matrix with positive entries in the diagonal.

The proof will be *constructive*, i.e.: we'll give a rule to build the Q and the R. Imagine that $a_1, \cdots, a_n$ are the columns of A (all of them vectors of $\mathbf{R}^m$). Then we can orthonormalize the set using the Gram-Schmidt procedure, and get $u_1, \cdots, u_n$, which make up an orthonormal basis for Col(A). But remember that, in order to obtain vector u_k in the Gram-Schmidt procedure, we need a_k, a_{k-1} and all others until a_1. But we don't need those vectors which come *after* a_k. Therefore, it is true that

$$a_k = r_{1k}u_1 + \cdots + r_{kk}u_k + 0 \cdot u_{k+1} + \cdots + 0 \cdot u_n$$

Let's call r_k that (column) vector of r's: $r_k = (r_{1k}, r_{2k}, \cdots, r_{kk}, 0, \cdots, 0)^T$. Now you can say that $a_k = Qr_k$, because you can say that a_k is a linear combination of the columns of Q, with weights given in r_k. Then:

$$A = [a_1 \cdots a_n] = [Qr_1 \cdots Qr_n] = QR$$

Where R is a matrix that has the r_k as columns. By construction, R is upper triangular, as we wanted.

Ufff... this was hard. Is this the way we're going to do it? By no means! In practice, we can start with Gram-Schmidt orthonormalization, for sure, but after that we realize that $Q^TQ = I$, since Q's columns are orthonormal. So, once Q is found,

$$Q^TA = Q^T(QR) = IR = R$$

So, as an example, we can find the QR factorization of

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

First of all, we apply Gram-Schmidt to the columns of A, obtaining Q. The final result is:

$$Q = \begin{pmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 2/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{pmatrix}$$

Now, we compute $R = Q^TA$ and get

$$R = \begin{pmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{pmatrix}$$

QR factorization is very often used in numerical calculations. But Gram-Schmidt, unfortunately, is very unstable for that, so a different technique is used to obtain it. Ufff, but that's another story which will be told somewhere else!

E12. Find the QR factorization for the following matrix:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$