
Three Short Theorems About Symmetric Matrices

0.1 Three Short Theorems About Symmetric Matrices

Matrices are incredibly useful for a multitude of reasons in programming and science in general. For one, a matrix can be used to store two dimensional collections of data, like tables or even images (which are collections of pixels). Arguably more importantly, matrices are synonymous with **linear maps**. A linear map is a special, “nicely behaved” function of one or more variables which we can do a lot of interesting math with, and in fact there is an entire branch of math which studies the ways in which we can approximate the world of non-linear functions by linear functions – otherwise known as calculus.

All three of these types of transformations are linear maps

I will assume that you know some basic facts about matrices and eigenvalues from now on. A suitable definition of symmetric matrices from first principles is that for any symmetric matrix A , $A = A^T$. In other words, A is symmetrical about the diagonal. It is often hard to interpret this abstract property about matrices in the physical or geometrical world, but it generally has something to do with symmetry. For example, it appears in the notion of “stress tensors” from physics, because of equal and opposite forces. And it appears in Hessian matrix from calculus, since partial derivatives can be done in any order.

The point is that symmetric matrices are useful, and you end up seeing them a lot. In this article, I’m going to be explaining 3 of my favorite facts about symmetric matrices in no particular order – as to how they’re all connected: who knows? This post is just a way for me to fit together 3 short relatively unrelated ideas into one post – think of it like a bargain deal!

Symmetric matrices have an orthonormal basis of eigenvectors This is often referred to as a “spectral theorem” in physics. We can define an **orthonormal** basis as a basis consisting only of unit vectors (vectors with magnitude 1) so that any two distinct vectors in the basis are perpendicular to one another (to put it another way, the inner product between any two vectors is 0). You might imagine such a basis is useful – because every such basis looks like the basis we always use in the real world (x, y, z) . Let’s start with proving this statement.

First, consider the following fact about dot products (inner products). We can represent them with matrix multiplication. In short, the dot product between two column vectors can be computed by making the first of those two a row vector:

$$x \cdot y = x^T y$$

This will be crucial to this short proof. Consider two distinct eigenvalues of A , λ_1 and λ_2 , with corresponding eigenvectors v_1 and v_2 . We will show that they are perpendicular to one another (once we have this, we can always pick multiple orthogonal vectors corresponding the same eigenvalue, and also we are free to pick vectors with length 1). We consider the following sum:

$$(Av_1)^T v_2 - v_1^T Av_2$$

We can simplify this a little bit knowing that v_1, v_2 are eigenvectors and thus $Av_1 = \lambda v_1$. We can rewrite (note that we can take constants out of matrix multiplication)

$$\begin{aligned} &= \lambda_1 v_1^T v_2 - \lambda_2 v_1^T v_2 \\ &= (\lambda_1 - \lambda_2)(v_1^T v_2) \\ &= (\lambda_1 - \lambda_2)(v_1 \cdot v_2) \end{aligned}$$

Now, we can use the fact that for any two matrices A and B , $AB^T = B^T A^T$. Rewriting the first equation above, we arrive at:

$$v_1^T A^T v_2 - v_1^T A v_2$$

But $A = A^T$, since A is symmetric! So in fact this quantity is zero. This means either $(\lambda_1 - \lambda_2)$ is zero (it isn't, since we picked two different eigenvalues); or $v_1 \cdot v_2$ is zero. Therefore, the two eigenvectors are perpendicular to one another.

Now, this lessens our workload considerably for a lot of purposes. Our eigenvectors automatically form a natural frame to work with.

Every square matrix can be written as the sum of a symmetric matrix and a skew-symmetric matrix. This is a fun, short proof. We define a **skew-symmetric** matrix as a matrix A where $A^T = -A$; so, reading the matrix horizontally or vertically returns the same matrix but with a flipped sign in each entry. We consider the following two sums:

$$M = \frac{1}{2}(A + A^T)$$

What can we say about this matrix? Well, we look at its entry in row i and column j (denoted M_{ij}), which is the sum $A_{ij} + A_{ji}$. But since addition is commutative, we also have

$$M_{ji} = M_{ij} = A_{ji} + A_{ij}$$

Indeed, the matrix reads the same horizontally and vertically; M is symmetric. In the same way, we can define:

$$N = \frac{1}{2}(A - A^T)$$

And by a nearly identical argument, we can show that N is skew symmetric. And looking at our definitions, we can see that $A = M + N$, and we are done.

Any square matrix is the product of two symmetric matrices In order to prove this statement, we have to cover the Jordan normal form, which is a unique way of representing every single matrix in a nice form.

You may remember from linear algebra the concept of diagonalization. Some matrices can be diagonalized, meaning they have a basis of eigenvectors. By applying a change of basis, then, we can arrive at the following equation:

$$M = SDS^{-1}$$

Where D is a diagonal matrix and S is a change of basis matrix whose columns are eigenvalues.

We can consider matrices as linear maps, which gives is the geometric intuition of diagonalization: for any diagonalizable matrix, we can pick a different set of axes where the transformation is simply scaling along each axis.

This transformation is just a scaling along the two marked axes.

It turns out that if you generalize this form, and take “almost diagonal” matrices instead of just diagonal ones, you end up with something called the Jordan normal form, which is unique for every matrix (basically, up to some reordering of “diagonal blocks”):

$$M = SJS^{-1}$$

Like any good generalization, it holds for the original; the Jordan normal form of a diagonalizable matrix is its diagonal matrix. In particular, symmetric matrices have the following nice Jordan form:

$$M = Q\Lambda Q^{-1}$$

Where Q is an orthogonal matrix, whose columns are the eigenvectors of M and therefore are: you guessed it, orthogonal!

The actual Jordan normal form looks like:

$$\begin{bmatrix} A_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & A_k \end{bmatrix}$$

Where each of the A_i refers to a **Jordan block**. For each eigenvalue λ of multiplicity m , we can have any number of associated Jordan blocks, whose sizes add up to m . Each Jordan block looks like:

$$\begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & \lambda \end{bmatrix}$$

The idea is that the matrix has eigenvalues along the diagonals, and 1s above the diagonals. We can use the following trick in our proof:

$$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & \lambda \\ 0 & \lambda & 1 \\ \lambda & 1 & 0 \end{bmatrix}$$

The key idea here is that for any non-diagonal Jordan block J , $J = AB$, where A is the matrix above with 1s on the off diagonals, and B is the remaining factor. The key thing here is that both A and B are symmetric. If our Jordan block is diagonal, on the other hand it is already symmetric. What we get from this is that any Jordan block can be written as the product of two symmetric matrices. You can probably see where we're going from here.

$$\begin{aligned}
 M &= SJS^{-1} \\
 &= S (AB) S^{-1} \\
 &= SA (S^T S^{T^{-1}}) BS^{-1} \\
 &= SAS^T (S^{T^{-1}} B^T S^{-1})
 \end{aligned}$$

The term on the left, however, is symmetric (check by taking its transpose; note that A is symmetric. By a similar argument, the right term is symmetric as well. So, we've proven that any square matrix can be written as the product of two symmetric matrices. In the real case, real square matrices are written as products of two real symmetric matrices.
