21 Symmetric and skew-symmetric matrices

21.1 Decomposition of a square matrix into symmetric and skew-symmetric matrices

Let $C_{n\times n}$ be a square matrix. We can write

$$C = (1/2)(C + C^t) + (1/2)(C - C^t) = A + B,$$

where $A^t = A$ is **symmetric** and $B^t = -B$ is **skew-symmetric**.

Examples:

- Let
$$C = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$
Then
$$C = (1/2)\left\{\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} + \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}\right\} + (1/2)\left\{\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} - \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}\right\},$$
and
$$C = (1/2)\begin{pmatrix} 2 & 6 & 10 \\ 6 & 10 & 14 \\ 10 & 14 & 18 \end{pmatrix} + (1/2)\begin{pmatrix} 0 & -2 & -4 \\ 2 & 0 & -2 \\ 4 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 5 \\ 3 & 5 & 7 \\ 5 & 7 & 9 \end{pmatrix} + \begin{pmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{pmatrix}.$$

- Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ be any differentiable function. Fix an $\mathbf{x}_0 \in \mathbb{R}^n$ and use Taylor's theorem to write
$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0) + D\mathbf{f}(\mathbf{x}_0)\Delta\mathbf{x} + \text{ higher order terms.}$$
Neglecting the higher order terms, we get what's called the first-order (or infinitessimal) approximation to $\mathbf{f}$ at $\mathbf{x}_0$. We can decompose the derivative $D\mathbf{f}(\mathbf{x}_0)$ into its symmetric and skew-symmetric parts, and write
$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{x}_0) + A(\mathbf{x}_0)\Delta\mathbf{x} + B(\mathbf{x}_0)\Delta\mathbf{x},$$
where $A = (1/2)(D\mathbf{f}(\mathbf{x}_0) + (D\mathbf{f}(\mathbf{x}_0)^t)$, and B is the difference of these two matrices.

This decomposition corresponds to the

> **Theorem of Helmholtz:** The most general motion of a sufficiently small non=rigid body can be represented as the sum of
> 1. A *translation* ($\mathbf{f}(\mathbf{x}_0)$)
> 2. A *rotation* (the skew-symmetric part of the derivative acting on $\Delta\mathbf{x}$), and
> 3. An *expansion (or contraction)* in three mutually orthogonal directions (the symmetric part).

Parts (2) and (3) of the theorem are not obvious; they are the subject of this chapter.

21.2 Skew-symmetric matrices and infinitessimal rotations

We want to indicate why a skew-symmetric matrix represents an infinitessimal rotation, or a "rotation to first order". Statements like this always mean: *write out a Taylor series expansion of the indicated thing, and look at the linear (first order) part.* Recall from the last chapter that a rotation in $\mathbb{R}^3$ is represented by an orthogonal matrix. Suppose we have a one-parameter family of rotations, say $R(s)$, where $R(0) = I$. For instance, we could fix a line in $\mathbb{R}^3$, and do a rotation through an angle s about the line. Then, using Taylor's theorem, we can write

$$R(s) = R(0) + (dR/ds)(0)s + \text{ higher order stuff.}$$

The matrix $dR/ds(0)$ is called an **infinitessimal rotation**.

Theorem: An infinitessimal rotation is skew-symmetric.

PROOF: As above, let $R(s)$ be a one-parameter family of rotations with $R(0) = I$. Then, since these are all orthogonal matrices, we have, for all s, $R^t(s)R(s) = I$. Take the derivative of both sides of the last equation:

$$d/ds(R^tR)(s) = dR^t(s)/dsR(s) + R^t(s)dR(s)/ds = 0,$$

since I is constant and $dI/ds = 0$ (the zero matrix). Now evaluate this at $s = 0$ to obtain

$$dR^t/ds(0)I + IdR/ds(0) = (dR/ds(0))^t + dR/ds(0) = 0.$$

If we write B for the matrix $dR/ds(0)$, this last equation just reads $B^t + B = 0$, or $B^t = -B$, so the theorem is proved.

Note: If you look back at Example 1, you can see that the skew-symmetric part of the 3×3 matrix has only 3 distinct entries: All the entries on the diagonal must vanish by skew-symmetry, and the $(1, 2)$ entry determines the $(2, 1)$ entry, etc. The three components above the diagonal, with a bit of fiddling, can be equated to the three components of a vector $\Psi \in \mathbb{R}^3$, called an *axial vector* since it's not really a vector. If this is done correctly, one can think of the direction of Ψ as the axis of rotation and the length of Ψ as the angle of rotation. You might encounter this idea in a course on mechanics.

21.3 Properties of symmetric matrices

For any square matrix A , we have

$$A\mathbf{x}\bullet\mathbf{y} = (A\mathbf{x})^t\mathbf{y} = \mathbf{x}^t A^t\mathbf{y} = \mathbf{x}\bullet A^t\mathbf{y}.$$

or

$$A\mathbf{x}\bullet\mathbf{y} = \mathbf{x}\bullet A^t\mathbf{y}.$$

In words, you can move A from one side of the dot product to the other by replacing A with A^t. Now suppose A is symmetric. Then this reads

$$A\mathbf{v}\bullet\mathbf{w} = \mathbf{v}\bullet A\mathbf{w}, \quad \forall \mathbf{v}, \mathbf{w} \ \in \mathbb{R}^n. \tag{1}$$

We'll need this result in what follows.

Theorem: The eigenvalues of a symmetric matrix are real numbers. The corresponding eigenvectors can always be assumed to be real.

Before getting to the proof, we need to review some facts about complex numbers:

> If $z = a + ib$ is complex, then its **real part** is a, and its **imaginary part** is b (both real and imaginary parts are real numbers). If $b = 0$, then we say that z is real; if $a = 0$, then z is imaginary. Its **complex conjugate** is the complex number $\bar{z} = a - ib$. A complex number is real $\iff$ it's equal to its complex conjugate: $\bar{z} = z$, (because this means that $ib = -ib$ which only happens if $b = 0$). The product of $\bar{z}$ with z is *positive* if $z \neq 0$: $\bar{z}z = (a - ib)(a + ib) = a^2 - (ib)^2 = a^2 - (i)^2b^2 = a^2 + b^2$. To remind ourselves that this is always ≥ 0, we write $\bar{z}z$ as $|z|^2$ and we call $\sqrt{|z|^2} = |z|$ the **norm** of z. For a complex vector $\mathbf{z} = (z_1, z_2, \ldots, z_n)^t$, we likewise have $\bar{\mathbf{z}}\bullet\mathbf{z} = |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2 > 0$, unless $\mathbf{z} = \mathbf{0}$.

Proof of the theorem: Suppose λ is an eigenvalue of the symmetric matrix A, and $\mathbf{z}$ is a corresponding eigenvector. Since λ might be complex, the vector $\mathbf{z}$ may likewise be a complex vector. We have

$$A\mathbf{z} = \lambda\mathbf{z}, \quad \text{and taking the complex conjugate of this equation,}$$

$$A\bar{\mathbf{z}} = \bar{\lambda}\bar{\mathbf{z}},$$

where we've used the fact that $\bar{A} = A$ since all the entries of A are real. Now take the dot product of both sides of this equation with $\mathbf{z}$ to obtain

$$A\bar{\mathbf{z}}\bullet\mathbf{z} = \bar{\lambda}\bar{\mathbf{z}}\bullet\mathbf{z}.$$

Now use (1) on the left hand side to obtain

$$A\bar{\mathbf{z}}\bullet\mathbf{z} = \bar{\mathbf{z}}\bullet A\mathbf{z} = \bar{\mathbf{z}}\bullet\lambda\mathbf{z} = \lambda\bar{\mathbf{z}}\bullet\mathbf{z}.$$

Comparing the right hand sides of this equation and the one above leads to

$$(\lambda - \bar{\lambda})\bar{\mathbf{z}}\bullet\mathbf{z} = 0. \tag{2}$$

Since $\mathbf{z}$ is an eigenvector, $\mathbf{z} \neq \mathbf{0}$ and thus, as we saw above, $\bar{\mathbf{z}}\bullet\mathbf{z} > 0$. In order for (2) to hold, we must therefore have $\lambda = \bar{\lambda}$, so λ is real, and this completes the first part of the proof. For the second, suppose $\mathbf{z}$ is an eigenvector. Since we now know that λ is real, when we take the complex conjugate of the equation

$$A\mathbf{z} = \lambda\mathbf{z},$$

we get

$$A\bar{\mathbf{z}} = \lambda\bar{\mathbf{z}}.$$

Adding these two equations gives

$$A(\mathbf{z} + \bar{\mathbf{z}}) = \lambda(\mathbf{z} + \bar{\mathbf{z}}).$$

Thus $\mathbf{z} + \bar{\mathbf{z}}$ is also an eigenvector corresponding to λ, and it's real. So we're done.

Comment: For the matrix

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix},$$

one of the eigenvalues is $\lambda = 1$, and an eigenvector is

$$\mathbf{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

But $(2 + 3i)\mathbf{v}$ is also an eigenvector, in principle. What the theorem says is that we can always find a real eigenvector. If A is real but not symmetric, and has the eigenvalue λ, then λ may well be complex, and in that case, there will not be a real eigenvector.

Theorem: The eigenspaces E_λ and E_μ are orthogonal if λ and μ are distinct eigenvalues of the symmetric matrix A.

Proof: Suppose $\mathbf{v}$ and $\mathbf{w}$ are eigenvectors of λ and μ respectively. Then

$$\begin{aligned} A\mathbf{v}\bullet\mathbf{w} &= \mathbf{v}\bullet A\mathbf{w} \quad (A \text{ is symmetric.) So} \\ \lambda\mathbf{v}\bullet\mathbf{w} &= \mathbf{v}\bullet\mu\mathbf{w} = \mu\mathbf{v}\bullet\mathbf{w}. \end{aligned}$$

This means that $(\lambda - \mu)\mathbf{v}\bullet\mathbf{w} = 0$. But $\lambda \neq \mu$ by our assumption. Therefore $\mathbf{v}\bullet\mathbf{w} = 0$ and the two eigenvectors must be orthogonal.

Example: Let

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Then $p_A(\lambda) = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1)$. Eigenvectors corresponding to $\lambda_1 = 3,\ \lambda_2 = -1$ are

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \text{and } \mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

They are clearly orthogonal, as advertised. Moreover, normalizing them, we get the *orthonormal basis* $\{\widehat{\mathbf{v}}_1, \widehat{\mathbf{v}}_2\}$. So the matrix

$$P = (\widehat{\mathbf{v}}_1 | \widehat{\mathbf{v}}_2)$$

is *orthogonal.* Changing to this new basis, we find

$$A_p = P^{-1}AP = P^tAP = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}.$$

In words: We have diagonalized the symmetric matrix A using an orthogonal matrix.

In general, if the symmetric matrix $A_{2\times 2}$ has distinct eigenvalues, then the corresponding eigenvectors are orthogonal and can be normalized to produce an o.n. basis. What about the case of repeated roots which caused trouble before? It turns out that everything is just fine *provided that A is symmetric.*

Exercises:

1. An arbitrary 2×2 symmetric matrix can be written as

   $$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

   where $a, b,$ and c are any real numbers. Show that $p_A(\lambda)$ has repeated roots if and only if $b = 0$ and $a = c$. (Use the quadratic formula.) Therefore, if $A_{2\times 2}$ is symmetric with repeated roots, $A = cI$ for some real number c. In particular, if the characteristic polynomial has repeated roots, then A is *already* diagonal. (This is more complicated in dimensions > 2.)

2. Show that if $A = cI$, and we use any orthogonal matrix P to change the basis, then in the new basis, $A_p = cI$. Is this true if A is just diagonal, but not equal to cI? Why or why not?

It would take us too far afield to prove it here, but the same result holds for $n \times n$ symmetric matrices as well. We state the result as a

Theorem: Let A be an $n \times n$ symmetric matrix. Then A can be diagonalized by an orthogonal matrix. Equivalently, $\mathbb{R}^n$ has an orthonormal basis consisting of eigenvectors of A.

Example (2) (cont'd): To return to the theorem of Helmholtz, we know that the skew-symmetric part of the derivative gives an infinitessimal rotation. The symmetric part $A = (1/2)(D\mathbf{f}(\mathbf{x}_0) + (D\mathbf{f})^t(\mathbf{x}_0))$, of the derivative is called the **strain tensor**. Since A is symmetric, we can find an o.n. basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ of eigenvectors of A. The eigenvectors determine the three *principle axes* of the strain tensor.

The simplest case to visualize is that in which all three of the eigenvalues are positive. Imagine a small sphere located at the point $\mathbf{x}_0$. When the elastic material is deformed by the forces, this sphere (a) will now be centered at $\mathbf{f}(\mathbf{x}_0)$, and (b) it will be rotated about some axis through its new center, and (c) this small sphere will be deformed into a small ellipsoid, the three semi-axes of which are aligned with the eigenvectors of A with the axis lengths determined by the eigenvalues.