Canonical forms

We saw that the solution of the initial value problem:

$$\frac{d\vec{v}}{dt} = A \cdot \vec{v}$$
$$\vec{v}(0) = \vec{v_0}$$

where A is a matrix with constant coefficients is given by:

$$\vec{v}(t) = \exp(tA) \cdot \vec{v_0}$$

We have also seen what $\exp(tA)$ looks like for some simple 2×2 matrices A.

Directly computing $\exp(tA)$ from A numerically is possible since the series converges rapidly. However, giving an algebraic formula would appear to be quite difficult since each computation of a matrix power is computationally intensive. The "canonical form" of the matrix A is a useful technique to solve this problem.

Conjugates of $\exp(tA)$

Given a (square) matrix A, we say that $G^{-1} \cdot A \cdot G$ is a conjugate of A, where G is an *invertible* matrix of the same size as A. By inspection of the power series term by term (which is enough due to absolute convergence!) we easily check that

$$G^{-1} \cdot \exp(tA) \cdot G = \exp(t(G^{-1} \cdot A \cdot G))$$

Note that the equation obtained by equating the coefficients of t^k is

$$G^{-1} \cdot A^k \cdot G = (G^{-1} \cdot A \cdot G)^k$$

which is easily checked by "multiplying out".

Recall that multiplying a vector by G amounts to a "linear change of co-ordinates". It is reasonably obvious that such a change of co-ordinates should not drastically change the behaviour of the solutions. Hence, it is natural to ask for the "simplest" form to which we can bring A by replacing it by $G^{-1} \cdot A \cdot G$.

The answer to this question is via the *Jordan canonical form*. For *any* matrix A (with coefficients lying within the field of complex numbers):

- A is of the form $S + N$ where S and N are linear combinations of $\mathbf{1}$ and the powers of A
- N is *nilpotent*; some power of N is 0
- S is *semi-simple*; it is diagonalisable over the field of complex numbers.

From the first statement it follows easily that $S \cdot N = N \cdot S$. Since we do not want to deal with complex numbers (measurements in science have to do with real numbers) we will employ a slightly different version of this result as given below.

Jordan Canonical Form for matrices over real numbers

For some chosen suitable order of the eigenvalues of the matrix S as above, let G' denote the change of co-ordinates to the basis consisting of eigenvectors. Then for a suitable choice of this order we can ensure that N consists mostly of 0's except possibly some 1's above the diagonal. A little bit of further algebraic manipulation using the fact that S is a matrix with real entries can be used to find a matrix G so that the following form holds.

- $G^{-1}AG$ takes the block diagonal form

$$G^{-1}AG = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_p \end{pmatrix}$$

- Further, the matrices A_k are square matrices that themselves have the block form

$$A_k = \begin{pmatrix} S_k & \mathbf{1}_{m_k} & 0 & \dots & 0 \\ 0 & S_k & \mathbf{1} & \dots & 0 \\ 0 & 0 & S_k & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & S_k \end{pmatrix}$$

  where S_k is a square matrix of size m_k. Note that it is the *same* square matrix that occurs in each block on the diagonal.
- The value of m_k is ether 1 or 2. In the first case S_k is just a 1×1 matrix, i.e. a scalar $S = (a_k)$. In the second case $(m_k = 2)$ we have

$$S_k = \begin{pmatrix} a_k & -b_k \\ b_k & a_k \end{pmatrix}$$

  for some real numbers a_k and b_k, where b_k is non-zero.
- Moreover, $G^{-1}SG$ is actually of the form

$$G^{-1}SG = \begin{pmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_p \end{pmatrix}$$

  where B_k is "like A_k but leave out the 1's above the diagonal". In other words,

$$B_k = \begin{pmatrix} S_k & 0 & 0 & \dots & 0 \\ 0 & S_k & 0 & \dots & 0 \\ 0 & 0 & S_k & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & S_k \end{pmatrix}$$

with S_k as before. It follows that $G^{-1}NG = G^{-1}AG - G^{-1}SG$ is the "remaining 1's" which makes a strictly upper triangular matrix which has some 1's above the diagonal, where there is at most one 1 in each row.

As seen above, the matrix G corresponds to a linear change of co-ordinates. Thus, to understand $\exp(tA)$ upto a linear change of co-ordinates, we need to understand $G^{-1}\exp(tA)G$. Since S and N commute, we see that $\exp(t(S+N)) = \exp(tS)\exp(tN)$. The term $\exp(tN)$ is rather simple as seen above, since $N^k = 0$ for some k, which gives a simple direct formula for $\exp(tN)$. To understand $G^{-1}\exp(tS)G$ we note that it has a block diagonal form with blocks made of $\exp(tS_k)$ for various values of k.

If $m_k = 1$, then S_k is a scalar 1×1 matrix and $\exp(tS_k) = (\exp(ta_k))$. In the cse $m_k = 2$ we use the matrix $I = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ introduced earlier to write $S_k = a_k\mathbf{1}_2 + b_kI$. Since I commutes with $\mathbf{1}_2$, it follows that

$$\exp(tS_k) = \exp(ta_k\mathbf{1}_2) \cdot \exp(tb_kI) = \exp(ta_k) \cdot \begin{pmatrix} \cos(tb_k) & -\sin(tb_k) \\ \sin(tb_k) & \cos(tb_k) \end{pmatrix}$$

In other words we have a scaling (or shrinking!) times a rotation.

In summary, the flow associated with $\exp(tA)$ is "made up" of the three types of flows that we studied earlier: scaling, rotation and shear.

Roots of the characteristic polynomial

Given a square matrix A, its characteristic polynomial is defined as $\det(t\mathbf{1} - A)$. Moreover, if G is an invertible matrix, then $\det(G^{-1}BG) = \det(B)$ for any matrix B. It follows that the characteristic polynomial of A is the same as the characteristic polynomial of $G^{-1}AG$. Since the latter is a block diagonal matrix for a suitable choice of G as above, we see that the characteristic polynomial of A is the product of the characteristic polynomials of A_k. One again exploiting the block form of A_k, one can show (though this is a little more difficult) that the characteristic polynomial is the M_k-th *power* of the characteristic polynomial of S_k where M_k is the number of row (or column) blocks in A_k; note that this is the same as the size of A_k divided by m_k.

It follows that any root of the characteristic polynomial of A is a root of $\det(t\mathbf{1}_1 - S_k)$ for a suitable k.

Examples

To make the above process clearer, let us study a few examples. Consider a general 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We need to find a suitable G to put A in canonical form.

First of all, the characteristic polynomial of A is $(t-a)(t-d)-bc$, or equivalently $t^2-(a+d)t+(ad-bc)$. This is a quadratic polynomial and we have three possibilities: - It is of the form $(t-p)(t-q)$ for *distinct* real numbers p and q - It is of the form $(t-p)^2$ - It is of the form $(t-p)^2+q^2$ for q a positive real number. When $t-p$ divides the characteristic polynomial, we see that $\det(p\mathbf{1}_2-A)=0$ and so there is a non-zero vector $\vec{v_p}$ (an *eigenvector* for A) such that $(p\mathbf{1}_2-A)\cdot\vec{v_p}=0$, or equivalently $A\cdot\vec{v_p}=p\vec{v_p}$. Hence, in the first case, we have a pair of eigen-vectors $\vec{v_p}$ and $\vec{v_q}$ associated with the roots $t=p$ and $t=q$ of the characteristic polynomials.

Exercise: Given distinct roots $p_1,\ldots,p_r$ of the characteristic polynomial, show that the associated eigenvectors $\vec{v_1},\ldots,\vec{v_r}$ are linearly independent.

As a consequence of this, we see that the column vectors $\vec{v_p},\vec{v_q}$ form a 2×2 invertible matrix G. We check easily that $G^{-1}AG=\begin{pmatrix}p & 0\\ 0 & q\end{pmatrix}$ which is in (block) diagonal form.

In the second case, we note that $(A-p\mathbf{1}_2)^2=0$ so that $A-p\mathbf{1}_2$ is a nilpotent matrix. In a suitable basis, we know that it is a strictly upper triangular matrix. In fact, in a suitable basis it is either the matrix of 0's or $\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}$. So, this is the change of basis that puts A in canonical form. Note that there *two* possible flows for this characteristic polynomial: the scaling by $\exp(tp)$ or the same scaling multiplied by a shearing operation.

In the last case, we consider the matrix $J=(A-p\mathbf{1}_2)/q$. This matrix satisfies $J^2=-\mathbf{1}_2$. It follows that for *any* non-zero vector $\vec{v}$, the vectors $\vec{v}$ and $J\cdot\vec{v}$ are linearly independent. In the change of basis given by these two vectors, one sees that the matrix of J becomes $\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$. In other words, the matrix of A in this basis becomes $\begin{pmatrix}p & -q\\ q & p\end{pmatrix}$ which is the required block form.

The case of matrices of larger size is certainly more complicated than that above. For example, we would need to have some way to *find* the roots of the characteristic polynomial. Moreover, having found them we also need to find the corresponding eigenvectors and so on. The theorem about Jordan canonical forms assures us that these steps can indeed be carried out.

As an example, we examine the case of a 4×4 matrix A. In this case, the characteristic polynomial has degree 4. The possible ways in which this splits up is:

- Two distinct quadratic polynomials $((t-p)^2+q^2)((t-u)^2+v^2)$ where p,q,u,v are real numbers and both q and v are positive.
- The square of a quadratic polynomial $((t-p)^2+q^2)^2$ where p,q are real numbers and q is positive.

- The product of a quadratic polynomial and two linear factors $((t-p)^2 + q^2)(t-u)(t-v)$ with p, q, u, v real numbers, q positive and $u \neq v$.
- The product of a quadratic polynomial and the square of a linear factor $((t-p)^2 + q^2)(t-u)^2$ with p, q, u real numbers and q positive.
- The product of 4 linear factors $(t-p)(t-q)(t-u)(t-v)$ with p, q, u, v distinct real numbers.
- The product of 2 linear factors and the square of a linear factor $(t-p)(t-q)(t-u)^2$ with p, q, u distinct real numbers.
- The product of a linear factor and the third power of a linear factor $(t-p)(t-q)^3$ with p, q distinct real numbers.
- The product the square of a linear factor and the square of another linear factor $(t-p)^2(t-q)^2$ with p, q distinct real numbers.
- The fourth power of a linear factor $(t-p)^4$.

In each case, the block structure of the semi-simple part S is determined by the number of *distinct* factors and the power of each distinct factor determines the number of repetitions of the associated block. If there is a repeated factor, then there *could* be a nilpotent block associated with it.