



Summary of week 3 (lectures 7, 8 and 9)

Lecture 7 commenced with an example similar to #7 on p. 41 of [VST]: solving a system of simultaneous linear differential equations. The technique applies to systems of the form

$$\underline{f}' = A\underline{f}$$

where $A \in \text{Mat}(n \times n, \mathbb{R})$ is given, \underline{f} is an unknown n -component column vector whose entries are differentiable functions $\mathbb{R} \rightarrow \mathbb{R}$, and \underline{f}' is the column vector obtained from \underline{f} by differentiating each component. The trick is to make a change of variables of the form

$$\underline{f} = T\underline{p}$$

form some appropriately chosen invertible $T \in \text{Mat}(n \times n, \mathbb{R})$. Doing this, the system becomes $T\underline{p}' = AT\underline{p}$, or

$$\underline{p}' = T^{-1}AT\underline{p}.$$

This new system of equations has the same form as the original, but the coefficient matrix has been changed from A to $T^{-1}AT$. If T can be chosen so that $T^{-1}AT$ is diagonal then a separation of variables is achieved, and the system becomes easy to solve.

The example done in the lecture was as follows:

$$\begin{aligned} f'(t) &= 2f(t) + 3g(t) \\ g'(t) &= 2f(t) + 7g(t) \end{aligned} \tag{1}$$

and so the solution process involves diagonalizing the matrix

$$A = \begin{pmatrix} 2 & 3 \\ 2 & 7 \end{pmatrix}.$$

It is assumed that you learnt how to do this in 1st year. The first step is to find the eigenvalues; these are the solutions of

$$\det \begin{pmatrix} 2 - \lambda & 3 \\ 2 & 7 - \lambda \end{pmatrix} = 0.$$

Expanding the determinant gives $(2 - \lambda)(7 - \lambda) - 6 = 0$, or

$$\lambda^2 - 9\lambda + 8 = 0.$$

The solutions are $\lambda = 8$ and $\lambda = 1$. These are the eigenvalues of A .

Next we find an eigenvector for each of the eigenvalues. These are nonzero vectors \underline{x} satisfying $(A - \lambda I)\underline{x} = \underline{0}$. Taking $\lambda = 8$ first, the equations to be solved become

$$\begin{pmatrix} -6 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The number of equations is the same as the number of unknowns; so one might expect that the only solution is the trivial solution $x = y = 0$. But the fact that the coefficient matrix has the form $A - \lambda I$ with λ an eigenvalue of A guarantees that the equations are redundant, and hence that a nontrivial solution exists. Indeed, in this case the first equation is clearly -3 times the second, and it is clear that

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{2}$$

is one nontrivial solution. Similarly, the eigenvalue 1 leads to the system

$$\begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

and one nontrivial solution is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}. \tag{3}$$

The vectors in Eq. (2) and Eq. (3) are the eigenvectors we sought. Note that any nonzero scalar multiple of an eigenvector is also an eigenvector, and so the vectors we have chosen could be replaced by nonzero scalar multiples of themselves. Different choices would be no better or worse than the choices we have made.

The change of variable matrix T should be chosen so that each of its columns is an eigenvector for A . It also needs to be invertible, so that $\underline{\tilde{f}} = T\underline{\tilde{p}}$ gives $\underline{\tilde{p}} = T^{-1}\underline{\tilde{f}}$ (so that $\underline{\tilde{f}}$ determines $\underline{\tilde{p}}$ and $\underline{\tilde{p}}$ determines $\underline{\tilde{f}}$). Unless T is invertible there is no guarantee that the transformed equations (involving $\underline{\tilde{p}}$) are equivalent to the original equations (involving $\underline{\tilde{f}}$). Fortunately, if the columns of T are chosen to be eigenvectors corresponding to distinct eigenvalues of A then it is guaranteed that T is invertible. This is a theorem that we shall prove later in the semester.

In accordance with the above remarks, we introduce new unknown functions p and q related to f and g via

$$\begin{pmatrix} f(t) \\ g(t) \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} p(t) \\ q(t) \end{pmatrix}.$$

In other words, in non-matrix notation,

$$\begin{aligned} f(t) &= p(t) + 3q(t), \\ g(t) &= 2p(t) - q(t). \end{aligned} \tag{4}$$

Differentiating (4) gives

$$\begin{aligned} f'(t) &= p'(t) + 3q'(t), \\ g'(t) &= 2p'(t) - q'(t). \end{aligned} \tag{5}$$

That is, we have $\tilde{f}' = T\tilde{p}'$ as well as $\tilde{f} = T\tilde{p}$. Since the first column of T is an 8-eigenvector of A and the second is a 1-eigenvector of A , it follows that $AT = TD$, where D is the diagonal matrix whose first diagonal entry is 8 and second diagonal entry is 1. You should check this by direct calculation:

$$\begin{pmatrix} -6 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 8 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6}$$

Using Eq. (4) and Eq. (5) to convert Eq. (1) into a system in the new unknowns p and q gives

$$\begin{pmatrix} p'(t) \\ q'(t) \end{pmatrix} = T^{-1}AT \begin{pmatrix} p(t) \\ q(t) \end{pmatrix} = \begin{pmatrix} 8 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p(t) \\ q(t) \end{pmatrix},$$

in view of Eq. (6). In non-matrix notation this becomes

$$\begin{aligned} p'(t) &= 8p(t) \\ q'(t) &= q(t) \end{aligned}$$

which can be regarded as two separate one-unknown systems instead of a single two-unknown system. This is a simplification, since for one-unknown systems the solutions are easy to find. For the equations above the solution above is

$$\begin{aligned} p(t) &= Ae^{8t} \\ q(t) &= Be^t \end{aligned}$$

where A and B are arbitrary constants. Using Eq. (4) to express this back in terms of the original unknowns, we see that the solution of the original system is

$$\begin{aligned} f(t) &= Ae^{8t} + 3Be^t \\ q(t) &= 2Ae^{8t} - Be^t \end{aligned}$$

where A and B are arbitrary.

The rest of Lecture 7, Lecture 8 and most of Lecture 9 dealt with §3d and §3e of [VST] (pp. 63–77), although Propositions 3.15, 3.21 and 3.22 have not yet been done in lectures. The key concept—and it is one of the most important concepts in the course—is that of a *subspace* of a vector space.

Given a vector space V it is natural to enquire whether it is possible for a subset of V to also be a vector space. In fact it is easy to find cases when this occurs. For example, we know that

$$\mathbb{R}^3 = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x, y, z \in \mathbb{R} \right\}$$

is a vector space over \mathbb{R} , with respect to the usual (componentwise) definitions of addition and scalar multiplication. Let S be the subset of \mathbb{R}^3 consisting of those triples for which x , y and z are all equal. That is,

$$S = \left\{ \begin{pmatrix} x \\ x \\ x \end{pmatrix} \mid x \in \mathbb{R} \right\}.$$

It is easily seen that S is also a vector space, addition and scalar multiplication being defined on S in the same way as for \mathbb{R}^3 itself. By contrast, the set T given by

$$T = \left\{ \begin{pmatrix} x \\ y \\ y+1 \end{pmatrix} \mid x, y \in \mathbb{R} \right\}$$

is not a vector space over \mathbb{R} , unless addition and scalar multiplication for T are defined in some other way, incompatible with the definitions used for \mathbb{R}^3 .

Suppose that V is a vector space over the field F , and that $U \subseteq V$ is also a vector space over F . If u, v are arbitrary elements of U then they are also elements of V , and so $u + v$ can be interpreted either as a sum of elements of U , using the addition operation for U , or as a sum of elements of V , using the addition operation for V . To say that these addition operations are compatible is to say that the answer is the same whichever of them is used. But by definition the sum of two elements of a vector space is also an element of that vector space; so we must have that $u + v \in U$. So it is a consequence of this compatibility requirement that applying V 's addition operation to elements of V that happen to lie in the subset U must give an element of U as the answer. Similarly, for the scalar multiplication operations on U and V to be compatible it is necessary that all scalar multiples of elements of V that lie in the subset U also lie in the subset U . Thus the subset U of V is closed under addition and scalar multiplication, in the sense of the following definition.

Definition. Let V be a vector space over the field F , and U a subset of V . Then

- (i) U is *closed under addition* if $u + v \in U$ for all $u, v \in U$;
- (ii) U is *closed under scalar multiplication* if $\lambda u \in U$ for all $u \in U$ and $\lambda \in F$.

It is clear that whenever $U \subseteq V$ is closed under addition then the addition operation on V gives rise to an addition operation on U . Similarly, if U is closed

under scalar multiplication then the scalar multiplication on V gives rise to a scalar multiplication on U . We say that U *inherits* addition and scalar multiplication operations from V .

Definition. Let V be a vector space over the field F . A *subspace* of V is a subset U of V that is closed under addition and scalar multiplication and is a vector space over F with respect to the addition and scalar multiplication operations that it inherits from V .

Examples

(1) Let S be the subset of \mathbb{R}^3 defined by

$$S = \left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} \mid a, b, c \in \mathbb{Z} \right\}$$

where \mathbb{Z} is the set of all integers. Then S is closed under addition but not closed under scalar multiplication.

(2) Let W be the subset of \mathbb{R}^3 defined by

$$W = \left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} \mid a, b, c \in \mathbb{R} \text{ and } a = \pm b \right\}.$$

Then W is closed under scalar multiplication but not addition.

(3) It is vacuously true that the empty subset of \mathbb{R}^3 is closed under addition and scalar multiplication. It is true that $\underline{u} + \underline{v} \in \emptyset$ whenever $\underline{u}, \underline{v} \in \emptyset$, simply because $\underline{u}, \underline{v} \in \emptyset$ is something that never happens!

If V is a vector space and $U \subseteq V$ is closed under addition and scalar multiplication then the fact that addition on V satisfies the associative law trivially implies that the inherited addition on U also satisfies the associative law. The fact that $\underline{u} + (\underline{v} + \underline{w}) = (\underline{u} + \underline{v}) + \underline{w}$ for all $\underline{u}, \underline{v}, \underline{w} \in V$ certainly implies that $\underline{u} + (\underline{v} + \underline{w}) = (\underline{u} + \underline{v}) + \underline{w}$ for all $\underline{u}, \underline{v}, \underline{w} \in U$, simply because all elements of U are elements of V . It is similarly obvious that the inherited operations satisfy most of the other vector space axioms as well. In fact, there are only two axioms to which this reasoning does not apply: the axiom that says that there is a the zero element and the axiom that says that each element has a negative. To guarantee that these axioms are also satisfied we need only assume in addition that the subset U is nonempty. It can then be shown that the zero element of V must be in the subset U , and that the negative of every element of U is in the subset U . It then follows readily that U satisfies all the vector space axioms.

Theorem. *If V is a vector space and U a subset of V that is nonempty and closed under addition and scalar multiplication, then U is a subspace of V .*

See p. 64 of [VST] for the details of the proof.

The above theorem often provides the most convenient way to prove that a given set is a vector space. Rather than going through all eight axioms and verifying them one at a time, it is often possible to show that the set in question is a subset of something that is already known to be a vector space (such as the set of all scalar valued functions on some set) and then show that it is nonempty and closed under addition and scalar multiplication.

For example, the set \mathcal{F} consisting of all real-valued functions on $[0, 1]$ is a vector space over \mathbb{R} , addition and scalar multiplication being defined as follows: if $f, g \in \mathcal{F}$ and $\lambda \in \mathbb{R}$ then $f + g, \lambda f \in \mathcal{F}$ are given by

$$\begin{aligned}(f + g)(t) &= f(t) + g(t) && \text{for all } t \in [0, 1] \\ (\lambda f)(t) &= \lambda f(t) && \text{for all } t \in [0, 1].\end{aligned}$$

Now let \mathcal{C} be the set consisting of all continuous functions $[0, 1] \rightarrow \mathbb{R}$. It is a standard result of calculus that the sum of two continuous functions is continuous; so \mathcal{C} is closed under addition. It is similarly a standard fact that a scalar multiple of a continuous function is continuous; so \mathcal{C} is closed under scalar multiplication. The zero element of \mathcal{F} is the function $z: [0, 1] \rightarrow \mathbb{R}$ defined by $z(t) = 0$ for all $t \in [0, 1]$. This is certainly continuous; so $z \in \mathcal{C}$, and hence $z \neq \emptyset$. Hence \mathcal{C} is a subspace of \mathcal{F} . In particular, \mathcal{C} is a vector space over \mathbb{R} .

For another example, let $\mathcal{S} = \{f \in \mathcal{C} \mid \int_0^1 f(t) dt = 0\}$. If z is the zero function (defined above) then $\int_0^1 z(t) dt = \int_0^1 0 dt = 0$, and so $z \in \mathcal{S}$. So $\mathcal{S} \neq \emptyset$. If $f, g \in \mathcal{S}$ and $\lambda \in \mathbb{R}$ then

$$\int_0^1 (f + g)(t) dt = \int_0^1 f(t) + g(t) dt = \int_0^1 f(t) dt + \int_0^1 g(t) dt = 0 + 0 = 0$$

and

$$\int_0^1 (\lambda f)(t) dt = \int_0^1 \lambda f(t) dt = \lambda \int_0^1 f(t) dt = \lambda 0 = 0$$

and so it follows that \mathcal{S} is closed under addition and scalar multiplication. Hence \mathcal{S} is a subspace of \mathcal{C} .

In this last example, the crucial thing that makes the proof work is that the expression $\int_0^1 f(t) dt$ depends linearly on f . To state this more precisely, the function $T: \mathcal{C} \rightarrow \mathbb{R}$ defined by $Tf = \int_0^1 f(t) dt$ is linear. In general, whenever we have a linear function from one vector space to another the set of all elements that are mapped to zero by the function constitutes a subspace of the domain of the function. This subspace is called the *kernel* of the linear function in question. It is also true that the image of the linear function is a subspace of the codomain. See [VST] pp. 66–69 for the proofs.

In Lecture 9 the following result was mentioned: if $A \in \text{Mat}(m \times n, F)$ and if $T: F^n \rightarrow F^m$ is defined by $T(v) = Av$ for all $v \in F^n$ then T is linear. This is proved in #11 on p. 59 of [VST]. (Note that there is a misprint in [VST] here: where it says “Exercise 3 of Chapter 1” it should say “Exercise 3 of Chapter 2”.) The kernel of this linear transformation is known as the *right null space* of the matrix A , denoted by $\text{RN}(A)$ (see Definition 7.24 of [VST]). Thus $\text{RN}(A)$ is the set of all solutions of the system of linear equations $Ax = 0$. As explained on p. 73 of [VST] the image of the function T is the *column space*, $\text{CS}(A)$, of the matrix A . By definition, this is the set of all linear combinations of the columns of A .

It is a general fact that if V is any vector space and v_1, v_2, \dots, v_k arbitrary elements of V then the set S consisting of all linear combinations of v_1, v_2, \dots, v_k is a subspace of V . We call S the *Span* of v_1, v_2, \dots, v_k :

$$S = \text{Span}(v_1, v_2, \dots, v_k) = \{ \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k \mid \lambda_1, \lambda_2, \dots, \lambda_k \in F \}.$$

The proof that S is a subspace is easy. Firstly, $0 = 0v_1 + 0v_2 + \dots + 0v_k \in S$; so $S \neq \emptyset$. Now suppose that $x, y \in S$. Then

$$\begin{aligned} x &= \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k \\ y &= \mu_1 v_1 + \mu_2 v_2 + \dots + \mu_k v_k \end{aligned}$$

for some scalars λ_i, μ_i , and we see that

$$x + y = (\lambda_1 + \mu_1)v_1 + (\lambda_2 + \mu_2)v_2 + \dots + (\lambda_k + \mu_k)v_k,$$

which is a linear combination of v_1, v_2, \dots, v_k , and hence an element of S . Thus S is closed under addition. Similarly, if $x \in S$ and λ is any scalar then

$$x = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k$$

for some scalars λ_i , and

$$\lambda x = (\lambda \lambda_1)v_1 + (\lambda \lambda_2)v_2 + \dots + (\lambda \lambda_k)v_k \in S.$$

So S is also closed under scalar multiplication, as required.

Example

Let A be the following matrix in $\text{Mat}(3 \times 4, \mathbb{R})$:

$$A = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 2 & -1 & -3 & 5 \\ 4 & 10 & 14 & 6 \end{pmatrix}.$$

Define $T: \mathbb{R}^4 \rightarrow \mathbb{R}^3$ by $T(\underline{x}) = A\underline{x}$ (for all $\underline{x} \in \mathbb{R}^4$). We shall find the right null space of A ; this amounts to solving the equations

$$\begin{aligned}x_1 + x_2 + x_3 + 2x_4 &= 0 \\2x_1 - x_2 - 3x_3 + 5x_4 &= 0 \\4x_1 + 10x_2 + 14x_3 + 6x_4 &= 0.\end{aligned}$$

The solution process should be familiar from last year: use row operations to reduce the system to an echelon form. Let us go through the steps anyway.

$$\begin{aligned}\begin{pmatrix} 1 & 1 & 1 & 2 \\ 2 & -1 & -3 & 5 \\ 4 & 10 & 14 & 6 \end{pmatrix} &\xrightarrow{\substack{R_2 := R_2 - 2R_1 \\ R_3 := R_3 - 4R_1}} \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -3 & -5 & 1 \\ 0 & 6 & 10 & -2 \end{pmatrix} \\ &\xrightarrow{R_3 := R_3 + 2R_2} \begin{pmatrix} \textcircled{1} & 1 & 1 & 2 \\ 0 & \textcircled{-3} & -5 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.\end{aligned}$$

When the echelon form is obtained, I recommend circling the leading entries of the nonzero rows, as indicated.

Note that the entries in the first column of the matrix correspond to the coefficients of x_1 in the various equations, the entries in the second column correspond to the coefficients of x_2 , and so on. The variables that correspond to columns containing no circled entries are free: this means that they can be given any values. The equations then determine the values of the other variables in terms of the values given to the free variables. Thus in the system above x_3 and x_4 are the free variables, and if we put $x_3 = s$ and $x_4 = t$ (arbitrary) then the second equation gives $x_2 = (-5/3)s + (1/3)t$, after which the first equation gives

$$x_1 = -x_2 - x_3 - 2x_4 = \frac{5}{3}s - \frac{1}{3}t - s - 2t = \frac{2}{3}s - \frac{7}{3}t.$$

So the general solution of the system is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \frac{2}{3}s - \frac{7}{3}t \\ -\frac{5}{3}s + \frac{1}{3}t \\ s \\ t \end{pmatrix} = s \begin{pmatrix} \frac{2}{3} \\ -\frac{5}{3} \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -\frac{7}{3} \\ \frac{1}{3} \\ 0 \\ 1 \end{pmatrix}$$

Thus we deduce that

$$\text{RN}(A) = \text{Span} \left(\left(\begin{pmatrix} \frac{2}{3} \\ -\frac{5}{3} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{7}{3} \\ \frac{1}{3} \\ 0 \\ 1 \end{pmatrix} \right) \right).$$

It is of course immediate from the definition that the column space of A is

$$\text{CS}(A) = \text{Span} \left(\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 10 \end{pmatrix}, \begin{pmatrix} 1 \\ -3 \\ 14 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \\ 6 \end{pmatrix} \right).$$

Finally, students should take particular notice of the following two definitions, which are of crucial importance in the theory. These are Definitions 3.17 and 3.18 of [VST].

Definition. Let V be a vector space over the field F .

- (i) Elements $v_1, v_2, \dots, v_k \in V$ are said to *span* V if $\text{Span}(v_1, v_2, \dots, v_k) = V$.
- (ii) Elements $v_1, v_2, \dots, v_k \in V$ are said to be *linearly independent* if the only solution of $\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k = \mathbf{0}$ for scalars $\lambda_1, \lambda_2, \dots, \lambda_k$ is given by $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$.

For example, suppose that $v_1, v_2, \dots, v_n \in F^m$ are columns of the $m \times n$ matrix A . Then

$$A \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n$$

and so to say that v_1, v_2, \dots, v_n are linearly independent is to say that the only solution of the system $Ax = \mathbf{0}$ is given by $x = \mathbf{0}$. In other words, the columns of the matrix A are linearly independent if and only if $\text{RN}(A) = \{\mathbf{0}\}$.