6. Vector Spaces and Modules

Before continuing, we review three examples from different branches of mathematics:

Example 1: Let's consider the solutions to the (second order, constant coefficient) ordinary differential equation

$$y'' - 5y' + 6y = 36x$$

The technique taught in introductory differential equations classes is to first find a single, particular solution[8] $y_p(x) = 6x + 5$ to the equation, and then find the general solution $y_h(x) = \alpha e^{2x} + \beta e^{3x 9}$ to the (homogeneous) equation

$$y'' - 5y' + 6y = 0$$

and add them together to find *all* the solutions to the original equation:

$$y(x) = y_p(x) + y_h(x) = (6x + 5) + \alpha e^{2x} + \beta e^{3x} \qquad \alpha, \beta \in \mathbb{R}$$

The constants α and β are then selected so that the solution satisfies any additional conditions we require (which we will omit).

This is very similar to solving $\mathbf{Ax} = \mathbf{b}$ when $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{A} \in M_{n\times m}(\mathbb{R})$. If you look back at the discussion of using RREF's to solve systems of equations, you will see that we: 1) found a particular solution to $\mathbf{Ax} = \mathbf{b}$, then 2) found *all* the solutions to $\mathbf{Ax} = \mathbf{0}$ (which vary with some collection of parameters), and 3) found *all* the solutions to $\mathbf{Ax} = \mathbf{b}$ by adding them together.

However, the similarities continue. One way of thinking about the system of linear equations is to see it in terms of functions. Define $T \colon \mathbb{R}^m \longrightarrow \mathbb{R}^n$ by $T(\mathbf{x}) = \mathbf{Ax}$. Then we are being asked to solve $T(\mathbf{x}) = \mathbf{b}$, given $\mathbf{b}$. For the differential equation, we let $C^k(\mathbb{R})$ be the set of all functions $f\, \mathbb{R} \longrightarrow \mathbb{R}$ where we can take derivatives k times and get a continuous function (so $C^0(\mathbb{R})$ is just the set of continuous functions). Then we can define $D \colon C^2(\mathbb{R}) \longrightarrow C^0(\mathbb{R})$ as

$$D(f) = \frac{d^2 f}{dx^2} - 5\frac{df}{dx} + 6f$$

And the differential equation is asking us to find all f such that $D(f) = 36x$.
With T and D in mind, notice that

(1) We can "add" elements of $\mathbb{R}^m$ and $C^2(\mathbb{R})$ and multiply them by real numbers. This is the usual vector addition and scalar multiplication in $\mathbb{R}^m$. In $C^2(\mathbb{R})$ it is the process of taking two functions f and g and defining $f + g$ to be the function which evaluates at x as $f(x) + g(x)$, or taking a real number α and a function f and setting $\alpha \cdot f$ to be the function $x \to \alpha \cdot \big(f(x)\big)$. For instance, we use these operations to form the functions $\alpha e^{2x} + \beta e^{3x}$ from the real numbers α, β and the functions e^{2x} and e^{3x}.

(2) The functions T and D interact well with the addition and multiplication:

$$\begin{aligned} T(\mathbf{x} + \mathbf{x}') &= \mathbf{A}(\mathbf{x} + \mathbf{x}') = \mathbf{Ax} + \mathbf{Ax}' = T(\mathbf{x}) + T(\mathbf{x}') \\ D(f + g) &= D(f) + D(g) \end{aligned}$$

due to the properties of matrix multiplication[10] and the fact that the derivative of a sum of functions is the sum of the derivatives.

[8] Using some method like undertermined coefficients or Laplace transforms

[9] Don't worry if you don't remember where this comes from – we won't need it

[10] which we will derive later, if you haven't seen this before

Example 2: Let $\mathcal{S}(\mathbb{F})$ be the set of all sequences of a field $\mathbb{F}$. That is, an element of $\mathcal{S}(\mathbb{F})$ is a ordered tuple

$$(s_1, s_2, s_3, \ldots) \qquad\qquad s_i \in \mathbb{F}, \forall i \in \mathbb{N}$$

(or, more succinctly, a function $s\colon \mathbb{N} \longrightarrow \mathbb{F}$ where we denote $s(i)$ by s_i). Just as above, we can add such sequences by add the i^{th} entry of one to the i^{th} entry of the other:

$$(s_1, s_2, s_3, \ldots) + (t_1, t_2, t_3, \ldots) = (s_1 + t_1, s_2 + t_2, s_3 + t_3, \ldots)$$

and multiply by an element $\lambda \in \mathbb{F}$ by multiplying each entry by λ:

$$\lambda \cdot (s_1, s_2, s_3, \ldots) = (\lambda \cdot s_1, \lambda \cdot s_2, \lambda \cdot s_3, \ldots)$$

We can define a function $P\colon \mathcal{S}(\mathbb{R}) \longrightarrow \mathcal{S}(\mathbb{R})$ by

$$P(s_1, s_2, s_3, \ldots) = (s_3 - 5s_2 + 6s_1, s_4 - 5s_3 + 6s_2, \ldots)$$

where the n^{th} entry of $P(s_1, s_2, s_3, \ldots)$ is $s_{n+2} - 5s_{n+1} + 6s_n$. Now consider the sequence

$$\mathbf{s} = (4, 8, 12, 16, \ldots) \in \mathcal{S} \qquad \Leftrightarrow \qquad s_n = 4n$$

What sequences $\mathbf{t} \in \mathcal{S}$ satisfy $P(\mathbf{t}) = \mathbf{s}$? Equating the n^{th} entries shows that this question is equivalent to solving the (constant coefficient) difference equation

$$s_{n+2} - 5s_{n+1} + 6s_n = 4n$$

for all $n \in \mathbb{N}$. By trying $s_n = a\,n + b$ we can obtain a single solution $s_n = 2n + 3$. That is, the sequence

$$(5, 7, 9, 11, 13, 15, \ldots)$$

is a particular solution. The theory of difference equations says that to obtain all the solutions, we should add to that any solution to the *homogeneous* equation $P(\mathbf{s}) = \mathbf{0} = (0, 0, 0, \ldots)$. The theory also provides a method for finding these solutions, which in this case gives sequences of the form $l_n = \alpha\, 2^n + \beta\, 3^n$. Thus, we obtain a solution for each $\alpha, \beta \in \mathbb{R}$ of the form $s_n = (2n+3) + \alpha\, 2^n + \beta\, 3^n$. Writing out the sequence gives the first few terms:

$$(5 + 2\alpha + 3\beta,\ 7 + 4\alpha + 9\beta,\ 9 + 8\alpha + 27\beta,\ 11 + 16\alpha + 81\beta,\ \ldots)$$

This obviously follows the same pattern. Does P have the same properties as T and D?

(Question 6.1) Verify that the map P has the same properties as T and D above, relative to $\mathcal{S}(\mathbb{F})$. That is, show that

$$P\big((s_1, s_2, s_3, \ldots) + (t_1, t_2, t_3, \ldots)\big) = P(s_1, s_2, s_3, \ldots) + P(t_1, t_2, t_3, \ldots)$$
$$P\big(\lambda \cdot (s_1, s_2, s_3, \ldots)\big) = \lambda \cdot P(s_1, s_2, s_3, \ldots)$$

for all sequences $(s_1, s_2, s_3, \ldots), (t_1, t_2, t_3, \ldots) \in \mathcal{S}(\mathbb{F})$.

Example 3: Let $\mathbb{Z}^3 = \big\{\, (a, b, c) \,\big|\, a, b, c \in \mathbb{Z} \,\big\}$. Then we can add elements in this set using

$$(a_1, b_1, c_1) + (a_2, b_2, c_3) = (a_1 + a_2, b_1 + b_2, c_1 + c_2)$$

and multiply *by integers* and stay in $\mathbb{Z}^3$: if $z \in \mathbb{Z}$ then

$$z \cdot (a, b, c) = (za, zb, zc) \in \mathbb{Z}^3$$

This set behaves in many respects like $\mathbb{R}^3$. Furthermore, if $\mathbf{A} \in M_{n \times m}(\mathbb{Z})$ then $\mathbf{x} \longrightarrow \mathbf{A}\mathbf{x}$ has the property that $\mathbf{A}(\mathbf{x} + \mathbf{y}) = \mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{y}$ and $\mathbf{A}(z\mathbf{x}) = z\mathbf{A}\mathbf{z}$. However, as we have seen, solving systems such as $\mathbf{A}\mathbf{x} = \mathbf{b}$ is trickier over $\mathbb{Z}$ since we can't always divide.

Nevertheless, we have three examples, all of which appear in both applied and pure mathematics, which have similar structures: we can add elements of a set, multiply them by elements of a field, and there are certain functions on them which respect the addition and multiplication. As before, we will extract the common structure and see what results follow after assuming that common structure to be present.

We want to be able to add elements of the set, and we have seen that it helps to require commutativity, associativity, an additive identity, and additive inverses. Since this is true in the examples above, we will include those as part of the structure. Similar conditions can be placed on the multiplication by elements of a ring. These conditions comprise the following definition:

Definition 23. *A* module *over a ring R is a tuple $(M, +, \cdot, \mathbf{0}_M)$ where $+ : M \times M \longrightarrow M$ and $\cdot : R \times M \longrightarrow M$ are functions, called addition and scalar multiplication, such that:*

1. *For all $\mathbf{x}, \mathbf{y} \in M$, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$*
2. *For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in M$, $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$*
3. *For all $\mathbf{x} \in M$, $\mathbf{x} + \mathbf{0}_V = \mathbf{x}$*
4. *For all $\mathbf{x} \in M$ there is an element $\mathbf{y} \in V$ such that $\mathbf{x} + \mathbf{y} = \mathbf{0}_V$*
5. *For all $\mathbf{x} \in M$ we have $1_R \cdot \mathbf{x} = \mathbf{x}$*
6. *For all $\alpha, \beta \in R$ and $\mathbf{x} \in M$, $(\alpha \cdot_R \beta) \cdot \mathbf{x} = \alpha \cdot (\beta \cdot \mathbf{x})$*
7. *For all $\alpha \in R$ and $\mathbf{x}, \mathbf{y} \in M$, $\alpha \cdot (\mathbf{x} + \mathbf{y}) = \alpha \cdot \mathbf{x} + \alpha \cdot \mathbf{y}$*
8. *For all $\alpha, \beta \in R$ and $\mathbf{x} \in M$, $(\alpha +_R \beta) \cdot \mathbf{x} = \alpha \cdot \mathbf{x} + \beta \cdot \mathbf{x}$*

$\mathbf{0}_M$ is called the zero element *in M.*

Notes:

1. Technically we have only allowed multiplication by elements of R on the left. We can ignore this when R is commutative, by taking $\mathbf{x} \cdot \alpha = \alpha \cdot \mathbf{x}$, and we will do this throughout these notes when R is commutative.
2. It *DOES* matter if R is non-commutative. Then the definition above is for a for a *left* module over $\mathbb{R}$, and then we must be careful not to multiply on the right. A similar definition describes the notion of a *right* module. It is possible (even for commutative rings) for M to be both a left and right module, but have essentially different definitions for left and right multiplication by elements of R.
3. When we need to identify the operations in M we will use subscripts: $+_M$. For instance in 8) above there is an addition from R (with the R subscript) and an addition in M, as well as scalar multiplication from M: $(a +_R b) \cdot_M \mathbf{x} = a \cdot_M \mathbf{x} +_M b \cdot_M \mathbf{x}$.

Because the addition in a module satisfies the four properties we identified as useful we *immediately* conclude that the vector $\mathbf{0}_M$ is unique and that the additive inverse of $\mathbf{x}$ is unique.

(Question 6.2) Explain why you already know the additive inverse of $\mathbf{x}$ equals $(-1) \cdot \mathbf{x}$. Don't give a proof!

(Question 6.3) Use the axioms to show that $0 \cdot \mathbf{x} = \mathbf{0}_M$. Note that this provides a means to identify $\mathbf{0}_M$ – just multiply any vector by $0 \in \mathbb{F}$.

Furthermore, just as associativity and commutativity of addition remove the ambiguities in how to add more that two elements, the additional properties using scalar multiplications imply that the expressions like

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \ldots + \alpha_k\mathbf{x}_k$$

(where $\alpha_1, \ldots, \alpha_k \in R$ are scalars, and $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are any elements in M) are unambiguous. We give expressions of this form a name:

Definition 24. *Let M be a module and let $S = [\mathbf{x}_1, \ldots, \mathbf{x}_k]$ be an ordered set of elements in M. An element $\mathbf{v} \in M$* linear combination *of the elements in S if there are some numbers $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that*

$$\mathbf{v} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \ldots + \alpha_k\mathbf{x}_k$$

Note:

(1) All of the α_i could be 0, so $\mathbf{0}$ is a linear combination of any other finite set of elements.
(2) We use an ordered set here to explicitly allow for repetitions and so that we can refer to a first, second, third, etc. element. However, using the properties of a module, any linear combination where some of the $\mathbf{x}_i$ are repeated can be reduced to a linear combination of distinct vectors.

In fact, this the most general type of expression we can construct by repeated applications of the addition and scalar multiplication. Using the rule in the definition, such an expression can always be reduced to a linear combination: for example, when $R = \mathbb{Z}$:

$$\begin{aligned}
3(4\mathbf{v}_1 - 2\mathbf{v}_2) + 2(\mathbf{v}_1 + 3\mathbf{v}_2) &= \big(3(4\mathbf{v}_1) + 3(-2\mathbf{v}_2)\big) + \big(2(\mathbf{v}_1) + 2(3\mathbf{v}_2)\big) && \big(\text{by } (7)\big) \\
&= \big((3 \cdot 4)\mathbf{v}_1 + (3 \cdot (-2))\mathbf{v}_2\big) + \big(2\mathbf{v}_1 + (2 \cdot 3)\mathbf{v}_2\big) && \big(\text{by } (6)\big) \\
&= \big(12\mathbf{v}_1 - 6\mathbf{v}_2\big) + \big(2\mathbf{v}_1 + 6\mathbf{v}_2\big) \\
&= \big(12\mathbf{v}_1 - 2\mathbf{v}_1\big) + \big(- 6\mathbf{v}_2 + 6\mathbf{v}_2\big) && \big(\text{by } (1), (2)\big) \\
&= (12 - 2)\mathbf{v}_1 + (-6 + 6)\mathbf{v}_2 && \big(\text{by } (8)\big) \\
&= 10\mathbf{v}_1 + 0\mathbf{v}_2 \\
&= 10\mathbf{v}_1 + \mathbf{0} && \big(\text{by an exercise}\big) \\
&= 10\mathbf{v}_1 && \big(\text{by } (3)\big)
\end{aligned}$$

In doing the following exercise, use only the properties 1)-8) in the definition of a module, and the properties of a ring. Keep track of each time you use a property and which one you use (one step at a time!).

(Question 6.4) Suppose $\mathbf{w}_1, \ldots, \mathbf{w}_j$ are linear combinations of $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$. Prove that any linear combination of $[\mathbf{w}_1, \ldots, \mathbf{w}_j]$ is also a linear combination of $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$.

When R is a field $\mathbb{F}$, modules go by a more familiar name:

Definition 25. *Let $\mathbb{F}$ be a field. An $\mathbb{F}$-vector space V is a module over $\mathbb{F}$. Elements of a vector space V are called* vectors

Just as fields admit simplifications due to the ability to divide, vector spaces will be the most nicely behaved modules.