ABSTRACT VECTOR SPACES AND THE CONCEPT OF ISOMORPHISM

MATH 196, SECTION 57 (VIPUL NAIK)

Corresponding material in the book: Sections 4.1 and 4.2.

Executive summary

General stuff ...

(1) There is an abstract definition of real vector space that involves a set with a binary operation playing the role of addition and another operation playing the role of scalar multiplication, satisfying a bunch of axioms. The goal is to axiomatize the key aspects of vector spaces.
(2) A subspace of an abstract vector space is a subset that contains the zero vector and is closed under addition and scalar multiplication.
(3) A linear transformation is a set map between two vector spaces that preserves addition and preserves scalar multiplication. It also sends zero to zero, but this follows from its preserving scalar multiplication.
(4) The *kernel* of a linear transformation is the subset of the domain comprising the vectors that map to zero. The kernel of a linear transformation is always a subspace.
(5) The *image* of a linear transformation is its range as a set map. The image is a subspace of the co-domain.
(6) The *dimension* of a vector space is defined as the size of any basis for it. The dimension provides an upper bound on the size of any linearly independent set in the vector space, with the upper bound attained (in the finite case) only if the linearly independent set is a basis. The dimension also provides a lower bound on the size of any spanning subset of the vector space, with the lower bound being attained (in the finite case) only if the spanning set is a basis.
(7) Every vector space has a particular subspace of interest: the zero subspace.
(8) The *rank* of a linear transformation is defined as the dimension of the image. The rank is the answer to the question: "how much survives the linear transformation?"
(9) The *nullity* of a linear transformation is defined as the dimension of the kernel. The nullity is the answer to the question: "how much gets killed under the linear transformation?"
(10) The sum of the rank and the nullity of a linear transformation equals the dimension of the domain. This fact is termed the *rank-nullity theorem*.
(11) We can define the *intersection* and *sum* of subspaces. These are again subspaces. The intersection of two subspaces is defined as the set of vectors that are present in both subspaces. The sum of two subspaces is defined as the set of vectors expressible as a sum of vectors, one in each subspace. The sum of two subspaces also equals the subspace spanned by their union.
(12) A linear transformation is *injective* if and only if its kernel is the zero subspace of the domain.
(13) A linear transformation is *surjective* if and only if its image is the whole co-domain.
(14) A *linear isomorphism* is a linear transformation that is *bijective*: it is both injective and surjective. In other words, its kernel is the zero subspace of the domain and its image is the whole co-domain.
(15) The dimension is an isomorphism-invariant. It is in fact a *complete isomorphism-invariant*: two real vector spaces are isomorphic if and only if they have the same dimension. Explicitly, we can use a bijection between a basis for one space and a basis for another. In particular, any n-dimensional space is isomorphic to $\mathbb{R}^n$. Thus, by studying the vector spaces $\mathbb{R}^n$, we have effectively studied all finite-dimensional vector spaces up to isomorphism.

Function spaces ...

(1) For any set S, consider the set $F(S, \mathbb{R})$ of *all* functions from S to $\mathbb{R}$. With pointwise addition and scalar multiplication of functions, this set is a vector space over $\mathbb{R}$. If S is finite (*not* our main case

of interest) this space has dimension $|S|$ and is indexed by a basis of S. We are usually interested in *subspaces* of this space.

(2) We can define vector spaces such as $\mathbb{R}[x]$ (the vector space of all polynomials in one variable with real coefficients) and $\mathbb{R}(x)$ (the vector space of all rational functions in one variable with real coefficients). These are both infinite-dimensional spaces. We can study various finite-dimensional subspaces of these. For instance, we can define P_n as the vector space of all polynomials of degree less than or equal to n. This is a vector space of dimension $n+1$ with basis given by the monomials $1, x, x^2, \dots, x^n$.

(3) There is a natural injective linear transformation $\mathbb{R}[x] \to F(\mathbb{R}, \mathbb{R})$.

(4) Denote by $C(\mathbb{R})$ or $C^0(\mathbb{R})$ the subspace of $F(\mathbb{R}, \mathbb{R})$ comprising the functions that are continuous everywhere. For k a positive integer, denote by $C^k(\mathbb{R})$ the subspace of $C(\mathbb{R})$ comrpising those functions that are at least k times continuously differentiable, and denote by $C^\infty(\mathbb{R})$ the subspace of $C(\mathbb{R})$ comprising all the functions that are *infinitely* differentiable. We have a descending chain of subspaces:

$$C^0(\mathbb{R}) \supseteq C^1(\mathbb{R}) \supseteq C^2(\mathbb{R}) \supseteq \dots$$

The image of $\mathbb{R}[x]$ inside $F(\mathbb{R}, \mathbb{R})$ lands inside $C^\infty(\mathbb{R})$.

(5) We can view differentiation as a linear transformation $C^1(\mathbb{R}) \to C(\mathbb{R})$. It sends each $C^k(\mathbb{R})$ to $C^{k-1}(\mathbb{R})$. It is surjective from $C^\infty(\mathbb{R})$ to $C^\infty(\mathbb{R})$. The kernel is constant functions, and the kernel of k-fold iteration is P_{k-1}. Differentiation sends $\mathbb{R}[x]$ to $\mathbb{R}[x]$ and is surjective to $\mathbb{R}[x]$.

(6) We can also define a formal differentiation operator $\mathbb{R}(x) \to \mathbb{R}(x)$. This is not surjective.

(7) Partial fractions theory can be formulated in terms of saying that some particular rational functions form a basis for certain finite-dimensional subspaces of the space of rational functions, and exhibiting a method to find the "coordinates" of a rational function in terms of this basis. The advantage of expressing in this basis is that the basis functions are particularly easy to integrate.

(8) We can define a vector space of sequences. This is a special type of function space where the domain is $\mathbb{N}$. In other words, it is the function space $F(\mathbb{N}, \mathbb{R})$.

(9) We can define a vector space of formal power series. The Taylor series operator and series summation operator are back-and-forth operators between this vector space (or an appropriate subspace therefore) and $C^\infty(\mathbb{R})$.

(10) Formal differentiation is a linear transformation $\mathbb{R}[[x]] \to \mathbb{R}[[x]]$. It is surjective but not injective. The kernel is the one-dimensional space of formal power series.

(11) We can consider linear differential operators from $C^\infty(\mathbb{R})$ to $C^\infty(\mathbb{R})$. These are obtained by combining the usual differentiation operator and multiplication operators using addition, multiplication (composition) and scalar multiplication. Finding the kernel of a linear differential operator is equivalent to solving a homogeneous linear differential equation. Finding the inverse image of a particular function under a linear differential operator amounts to solving a non-homogeneous linear differential equation, and the solution set here is a translate of the kernel (the corresponding solution in the homogeneous case, also called the *auxilliary solution*) by a particular solution. The first-order case is particularly illuminative because we have an explicit formula for the fibers.

1. Abstract definitions

1.1. **The abstract definition of a vector space.** A *real vector space* (just called *vector space* for short) is a set V equipped with the following structures:

- A binary operation $+$ on V called addition that is commutative and associative. By "binary operation" we mean that it is a map $V \times V \to V$, i.e., it takes two inputs in V and gives an output in V. Explicitly, for any vectors $\vec{v}, \vec{w}$ in V, there is a vector $\vec{v} + \vec{w} \in V$. The operation is commutative and associative:
  - *Commutativity* means that for any vectors $\vec{v}, \vec{w} \in V$, $\vec{v} + \vec{w} = \vec{w} + \vec{v}$.
  - *Associativity* means that for any vectors $\vec{u}, \vec{v}, \vec{w} \in V$, $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.
- A special element $\vec{0} \in V$ that is an identity for addition. Explicitly, for any vector $\vec{v} \in V$, we have $\vec{0} + \vec{v} = \vec{v} + \vec{0} = \vec{v}$.
- A scalar multiplication operation $\mathbb{R} \times V \to V$ denoted by concatenation such that:

- $0\vec{v} = \vec{0}$ (the $\vec{0}$ on the right side being the vector 0) for all $\vec{v} \in V$.
- $1\vec{v} = \vec{v}$ for all $\vec{v} \in V$.
- $a(b\vec{v}) = (ab)\vec{v}$ for all $a, b \in \mathbb{R}$ and $\vec{v} \in V$.
- $a(\vec{v} + \vec{w}) = a\vec{v} + a\vec{w}$ for all $a \in \mathbb{R}$ and $\vec{v}, \vec{w} \in V$.
- $(a + b)\vec{v} = a\vec{v} + b\vec{v}$ for all $a, b \in \mathbb{R}$, $\vec{v} \in V$.

Note that the vector spaces $\mathbb{R}^n$ that we have encountered are examples of real vector spaces in the sense above. However, there are many other vector spaces, such as spaces of functions, that at least superficially look very different.

1.2. **Abstract vector spaces: where do they live?** One of the difficulties that many people have with grasping abstract vector spaces is that it's not clear *where* this vector space is. With $\mathbb{R}^n$, we know what the elements (vectors) are: they are vectors with n coordinates, all of which are real numbers. We know the algebraic representation, and also have a vague geometric picture. Admittedly, the geometric picture is clear only for $n = 1, 2, 3$, but we can obtain our intuition from there and extend formally from there.

With abstract vector spaces, *where* they live could vary quite a bit based on what space we are considering. But in some sense, it doesn't matter *where they live*. What really matters is how the vectors interact with each other, i.e., how they add and scalar multiply. The addition and scalar multiplication are the essence of a vector space. How the vectors are written or what they are called is less relevant than how they add. Just like where you live or where you were born is not directly relevant to your grade: how you score on the test is. We judge vectors not by how they're written, but by the way they add and scalar multiply. We'll understand this better when we study the definition of *isomorphism* of vector spaces.

1.3. **What are all those axioms for?** The conditions such as commutativity, associativity, distributivity, etc. imposed in the abstract definition of vector space are there in order to make sure that the *key features* of the concrete vector spaces we have encountered so far are preserved in the abstract setting. Basically, what we want is that any algebraic identity or manipulation technique that we need to use in our usual proofs is available to us in the abstract setting.

1.4. **The abstract definition of subspace and linear transformation.** Fortunately, these definitions don't really differ from the definitions you are probably already familiar with from earlier. The reason is that it's only the beginning part (the foundation, so to speak) that gets more complicated in the abstract setting. The rest of it was already sufficiently abstract to begin with. Nonetheless, we review the definitions below.

A *(linear) subspace* of a vector space is defined as a nonempty subset that is closed under addition and scalar multiplication. In particular, any subspace must contain the zero vector. So, an alternative definition of subspace is that it is a subset that contains the zero vector, is closed under addition, and is closed under scalar multiplication. Note that if we just say *subspace* we are by default referring to a linear subspace.

A subspace of a vector space can be viewed as being a vector space in its own right. Note that there is one vector that we are sure every subspace must contain: the zero vector.

Suppose V and W are vector spaces. A function $T : V \to W$ is termed a *linear transformation* if T preserves addition and scalar multiplication, i.e., we have the following two conditions:

- $T(\vec{v_1} + \vec{v_2}) = T(\vec{v_1}) + T(\vec{v_2})$ for all vectors $\vec{v_1}, \vec{v_2} \in V$.
- $T(a\vec{v}) = aT(\vec{v})$ for all $a \in \mathbb{R}$, $\vec{v} \in V$.

Note that any linear transformation must send the zero vector to the zero vector. This need not be imposed as a separate condition: it follows from the scalar multiplication condition.

1.5. **Kernel and image.** The *kernel* of a linear transformation $T : V \to W$ is defined as the set of all vectors $\vec{v} \in V$ such that $T(\vec{v})$ is the zero vector. As we saw earlier, the kernel of any linear transformation is a *subspace* of V. In other words, it is non-empty (note that in particular it contains the zero vector of V), it is closed under addition, and it is closed under scalar multiplication.

The *image* of a linear transformation $T : V \to W$ is defined as the set of all vectors $\vec{w} \in W$ that can be written as $\vec{w} = T(\vec{v})$ for some vector $\vec{v} \in V$. In the language of functions, it is simply the range of T. The image of T is a *subspace* of W.

The proofs of both statements (the kernel is a subspace and the image is a subspace) are the same as those we saw earlier when introducing the concept of kernel. However, now that we are dealing with abstract vector spaces as opposed to the concrete setting, we need to be sure that every manipulation that we perform is included in, or justifiable from, the axioms in the abstract setting.

1.6. **Dimension and containment.** Suppose V is a real vector space. The *dimension* of V (as a real vector space) is defined in the following equivalent ways:

(1) It is the maximum possible size of a linearly independent set in V. Note that in the finite case, a linearly independent set has this maximum size if and only if it is a basis.
(2) It the size of any basis of V.
(3) It is the minimum possible size of a spanning set in V. Note that in the finite case, a spanning set has this minimum size if and only if it is a basis.

We call a vector space *finite-dimensional* if its dimension is finite, and *infinite-dimensional* otherwise.

The following are true for a subspace containment: suppose U is a subspace of a real vector space V. Then, the dimension of U is less than or equal to the dimension of V. If U is finite-dimensional, then the dimensions are equal if and only if $U = V$.

1.7. **The zero subspace.** Every vector space has a particular subspace of interest: the *zero subspace*. This is the subspace that contains only the zero vector. This is a zero-dimensional space. In other words, the empty set is a basis for it.

1.8. **Rank-nullity theorem.** The rank-nullity theorem holds for abstract vector spaces. Suppose $T : V \to W$ is a linear transformation from a real vector space V to a real vector space W. Suppose further that V is finite-dimensional. We do not need to assume anything regarding whether W is finite-dimensional.

Recall that the *rank* of T is defined as the dimension of the image of T. The *nullity* of T is defined as the dimension of the kernel of T. The rank-nullity theorem states that the sum of the rank of T and the nullity of T is the dimension of the domain space V.

This is the same as the old rank-nullity theorem, except that now, we are no longer thinking of things in terms of matrices, but simply in terms of abstract spaces and linear transformations between them.

1.9. **Sum and intersection of subspaces.** We had earlier defined the concepts of sum and intersection of subspaces. The same concepts apply with the same definition in the abstract setting. Explicitly:

- If U_1 and U_2 are subspaces of a real vector space V, then the intersection $U_1 \cap U_2$, defined as the set of vectors that are in both U_1 and U_2, is also a subspace of V.
- If U_1 and U_2 are subspaces of a real vector space V, then the sum $U_1 + U_2$, defined as the set of vectors that can be expressed as the sum of a vector in U_1 and a vector in U_2, is also a subspace of V.

  Note that the union $U_1 \cup U_2$ is just a subset and not in general a subspace, and in fact, $U_1 \cup U_2 \subseteq U_1 + U_2$ and $U_1 + U_2$ is the subspace spanned by $U_1 \cup U_2$. It is a subspace if and only if either $U_1 \subseteq U_2$ or $U_2 \subseteq U_1$, and further, that happens if and only if $U_1 \cup U_2 = U_1 + U_2$.

2. Isomorphism of vector spaces

2.1. **Definition of isomorphism.** Recall our general concept of *isomorphism* from earlier in the course: it is an invertible mapping that preserves the essence of the structure. In the context of vector spaces, a *linear isomorphism* between abstract vector spaces V and W is a bijective linear transformation $T : V \to W$.

If V and W are vector spaces and there exists a linear isomorphism $T : V \to W$, then we say that V and W are isomorphic.

2.2. **Isomorphism as an equivalence relation.** Isomorphism is an equivalence relation between vector spaces. Explicitly, it satisfies the following three conditions:

- *Reflexivity*: Every vector space is isomorphic to itself. We can choose the identity map as the isomorphism (though it is not the only possible isomorphism).
- *Symmetry*: If V and W are isomorphic, then W and V are isomorphic. Explicitly, if $T : V \to W$ is a linear isomorphism, then $T^{-1} : W \to V$ is a linear isomorphism.

- *Transivity*: If U and V are isomorphic, and V and W are isomorphic, then U and W are isomorphic. Explicitly, if $T_1 : U \to V$ is a linear isomorphism and $T_2 : V \to W$ is a linear isomorphism, then $T_2 \circ T_1 : U \to W$ is a linear isomorphism. Explicitly, $(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$.

2.3. **Isomorphism and dimension.** The dimension of a vector space is an invariant that completely determines the isomorphism class. Explicitly, if V and W are vector spaces, then V and W are isomorphic if and only if they have the same dimension. Constructively, the isomorphism is obtained as follows: choose any set bijection from a basis of V to a basis of W, and then extend linearly to a linear isomorphism from V to W.

This is particularly useful for finite-dimensional vector spaces: given a n-dimensional vector space and a m-dimensional vector space, the vector spaces are isomorphic if and only if $n = m$. In particular, this also tells us that any n-dimensional real vector space is isomorphic to the "standard" n-dimensional vector space $\mathbb{R}^n$. Put another way, studying the spaces $\mathbb{R}^n, n \in \mathbb{N}_0$ is tantamount to studying all finite-dimensional vector spaces up to isomorphism. That's why our concreteness so far didn't really lose us much generality.

The particular case of dimension zero gives the zero space. This is isomorphic to the zero subspace in any vector space.

3. Function spaces

3.1. **The general idea of function spaces.** The idea of *function spaces* is as follows. For S any set, we can define the space of *all* functions from S to $\mathbb{R}$ and make this a real vector space with the following structure:

- The addition of functions is defined *pointwise*. Explicitly, given functions $f, g : S \to \mathbb{R}$, we define $f + g$ as the function $x \mapsto f(x) + g(x)$.
- Scalar multiplication is defined as follows: given $\lambda \in \mathbb{R}$ and $f : S \to \mathbb{R}$, λf is defined as function $x \mapsto \lambda f(x)$.

We will denote this vector space as $F(S, \mathbb{R})$.

Now, this is the space of *all* functions on the set S. We are usually interested in other vector spaces that arise as subspaces of this space. Specifically, we are interested in subsets of this space that contain the zero function (the function sending everything to zero), are closed under pointwise addition of functions, and are closed under scalar multiplication.

Note also that the "vectors" here are now "functions." This requires a bit of rethinking, because we are used to thinking of vectors as pointy arrows or equivalently as things with coordinates. Functions, on the other hand, do not look like that. But to make a vector space, we don't care about whether the things look like our preconceived notion of vectors, but rather, we care about whether they have the addition and scalar multiplication operations satisfying the conditions we have specified.

3.2. **A basis for the space of all functions when the set is finite.** If S is finite, then the space of all functions on a set S has a basis indexed by the elements of S. For each $s \in S$, define the characteristic function $\mathbf{1}_s$ as the function that sends s to 1 and sends all other elements of S to 0. For any function f, the expression for f in terms of this basis is:

$$f = \sum_{s \in S} f(s)\mathbf{1}_s$$

Note that this idea cannot be used when S is infinite because the required sum would be infinite, and vector spaces only permit finite sums.

3.3. **The vector space of polynomials.** The set of *all* polynomials with real coefficients in one variable x is a vector space, with the usual definition of addition and scalar multiplication of polynomials. This vector space is sometimes denoted $\mathbb{R}[x]$ (the book denotes this space by P). Note that there is *also* a definition of *multiplication* of polynomials but that definition is *not* part of the vector space structure.

Explicitly, an element of this vector space is of the form:

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

Now, any polynomial can be thought of as a function. In other words, given any polynomial p, we can think of the *function* $x \mapsto p(x)$. In other words, we have a mapping:

$$\mathbb{R}[x] \to F(\mathbb{R}, \mathbb{R})$$

that sends any polynomial to the corresponding function. We note the following about the mapping:

- The mapping is *linear*: It preserves sums and scalar multiples. What this means is that adding polynomials as *polynomials* and then considering them as functions is equivalent to considering them as functions and then adding as functions. This is no surprise: our formal method for polynomial addition is designed to mimic function addition. Also, scalar multiplying a polynomial and *then* viewing it as a function is equivalent to converting it to a function and then scalar multiplying that function.
- The mapping is *injective*: What this means is that two different polynomials can never give rise to the same function. Equivalently (since the mapping is linear) no nonzero polynomial can be the zero function. This is obvious: a nonzero polynomial of degree n has at most n real roots, hence cannot be zero at all points.

Thus, the image of $\mathbb{R}[x]$ is a subspace of $F(\mathbb{R}, \mathbb{R})$. Since the mapping is injective, we can think of this subspace as another copy of $\mathbb{R}[x]$, so we sometimes abuse notation and identify $\mathbb{R}[x]$ with that image.

3.4. **Subspaces of the space of polynomials.** The space $\mathbb{R}[x]$ of polynomials is infinite-dimensional. The following is the most convenient basis for it:

$$1, x, x^2, x^3, \dots$$

Note that this is a spanning set because every polynomial is expressible as a linear combination of a finite number of these. It is a basis because there are no linear relations between these.

We can define the following subspaces of $\mathbb{R}[x]$ of interest. For any nonnegative integer n, let P_n be the span of the subset:

$$1, x, x^2, \dots, x^n$$

In other words, P_n is the $(n+1)$-dimensional space comprising all polynomials of degree $\leq n$. The above set forms a basis for P_n.

We have the following subspace inclusions:

$$P_0 \subseteq P_1 \subseteq P_2 \subseteq P_3 \subseteq \dots$$

And the whole space is the union of all these subspaces, i.e., we have:

$$\mathbb{R}[x] = P = \bigcup_{i=0}^{\infty} P_i$$

3.5. **The vector space of rational functions.** We denote by $\mathbb{R}(x)$ the set of all rational functions, i.e., expressions of the form $p(x)/q(x)$ where p and q are both polynomials with q not the zero polynomial, up to the following *equivalence* relation: $p_1(x)/q_1(x)$ and $p_2(x)/q_2(x)$ are considered the "same" rational function if $p_1(x)q_2(x) = p_2(x)q_1(x)$.

With the usual definition of addition (take common denominator, then add numerators) and scalar multiplication (just multiply the scalar in the numerator), the space of rational functions is a vector space. Further, there is a natural injective linear transformation from the space of polynomials to the space of rational functions:

$$\mathbb{R}[x] \to \mathbb{R}(x)$$

that sends a polynomial $p(x)$ to $p(x)/1$.

The map is not surjective, because there do exist rational functions that are not polynomials.

We might be tempted to say that there is a natural map:

$$\mathbb{R}(x) \to F(\mathbb{R}, \mathbb{R})$$

However, this would be inaccurate, because the *function* defined by a rational function is not defined at the points where the denominator becomes zero. So, the above map does not *quite* make sense. There are ways of getting around the issue by fixing either the domain or the co-domain appropriately, but we shall not bother right now.

4. Evaluation functionals and related transformations

4.1. Evaluation functional at a single point. For any real number u, the evaluation functional eval_u is defined as a linear transformation:

$$\text{eval}_u : F(\mathbb{R}, \mathbb{R}) \to \mathbb{R}$$

given by:

$$\text{eval}_u(f) := f(u)$$

The term *linear functional* is used for a linear transformation from a vector space to the vector space $\mathbb{R}$ (viewed as a one-dimensional vector space over itself). The evaluation maps are linear functionals.

4.2. Evaluation at multiple points simultaneously. Consider a tuple $(u_1, u_2, \dots, u_n)$ of real numbers. We can define a linear transformation:

$$\text{eval}_{(u_1, u_2, \dots, u_n)} : F(\mathbb{R}, \mathbb{R}) \to \mathbb{R}^n$$

as follows:

$$f \mapsto \begin{bmatrix} f(u_1) \\ f(u_2) \\ \cdot \\ \cdot \\ \cdot \\ f(u_n) \end{bmatrix}$$

This linear transformation involves the simultaneous evaluation of f at multiple points, and it further involves storing the outputs as the coordinates of a vector.

4.3. Evaluation transformations from smaller spaces. Instead of considering evaluation transformations originating from $F(\mathbb{R}, \mathbb{R})$, we can consider evaluation transformations originating from smaller spaces. For instance, recall that we defined P_m as the vector space of polynomials of degree $\leq m$. This is a $(m+1)$-dimensional real vector space. Given n distinct points, we can define the evaluation transformation:

$$P_m \to \mathbb{R}^n$$

4.4. Injectivity and surjectivity of the evaluation transformation. The following are true:

- The evaluation transformation from a function space to $\mathbb{R}^n$ (based on evaluation at a collection of points) is *injective* if and only if the only function that evaluates to zero at all the points in that collection is the zero function.
- The evaluation transformation from a function space to $\mathbb{R}^n$ (based on evaluation at a collection of points) is *surjective* if and only if every possible tuple of output values at that collection of points arises from a function in that function space.

4.5. **Setting things up using matrices: need for choosing a basis of the space, and hence parameters.** Consider in more detail the evaluation transformation:

$$P_m \to \mathbb{R}^n$$

Note that P_m, the space of polynomials of degree $\le m$, is an abstract vector space. Although it has dimension $(m+1)$ we do not think of it as being the same as $\mathbb{R}^{m+1}$. If we choose a basis for P_m, then we can write coordinates in that basis, and we can then think of the map as being like a map $\mathbb{R}^{m+1} \to \mathbb{R}^n$, and describe it with a $n \times (m+1)$ matrix.

The obvious choice of basis is:

$$1, x, x^2, \dots, x^m$$

Thus, for a polynomial:

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$$

the corresponding coordinates are:

$$\begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{bmatrix}$$

Recall that, at the start of the course, we had talked about functional forms that are *linear in the parameters*. We can really think about such functional forms as describing vector spaces of the functions, with the functions appearing in front of the parameters as a basis for that function space, and the parameters specifying the coordinates of a particular function of the function space in terms of that basis. For the example of P_m, the functional form of a polynomial of degree $\le m$ corresponds to the function space P_m. This space has basis $1, x, x^2, \dots, x^m$ and the coefficients that appear in front of these monomials in the description of a polynomial are the coordinates in that basis. These coordinates are our parameters.

Explicitly, the dictionary between our earlier jargon and our new jargon in as follows:

- Parameters for the general functional form $\leftrightarrow$ Coordinates in the chosen basis for the function space
- Inputs (of input-output pair fame) $\leftrightarrow$ Points at which we are performing the evaluation maps
- Outputs $\leftrightarrow$ Outputs of the evaluation maps

5. Function spaces: differentiation and integration

5.1. **Spaces of continuous and differentiable functions.** We denote by $C(\mathbb{R})$ or $C^0(\mathbb{R})$ the subset of $F(\mathbb{R}, \mathbb{R})$ comprising all the functions that are continuous on all of $\mathbb{R}$. This is a subspace. Here's why:

- The zero function is continuous.
- A sum of continuous functions is continuous: Remember that this follows from the fact that the limit of the sum is the sum of the limits, which in turn can be proved using the $\varepsilon - \delta$ definition of the limit (there are also alternative ways of thinking about it).
- A scalar multiple of a continuous function is continuous: This follows from the fact that the limit of a scalar multiple is the corresponding scalar multiple of the limit.

The elements ("vectors") of the vector space $C(\mathbb{R})$ are continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Note that this space is pretty huge, but relative to $F(\mathbb{R}, \mathbb{R})$, it is still quite small: if you just picked a function with random values everywhere, it would probably be *very far* from continuous. In fact, it's unlikely to be continuous at *any* point.

The space $C(\mathbb{R})$, also denoted $C^0(\mathbb{R})$, has a number of interesting subspaces. For k a positive integer, we define $C^k(\mathbb{R})$ as the subspace of $C(\mathbb{R})$ comprising those continuous functions that are at least k times continuously differentiable on all of $\mathbb{R}$. Explicitly, $f \in C^k(\mathbb{R})$ if $f^{(k)}$ exists and is in $C(\mathbb{R})$. We thus have subspace inclusions:

$$C(\mathbb{R}) = C^0(\mathbb{R}) \supseteq C^1(\mathbb{R}) \supseteq C^2(\mathbb{R}) \supseteq \dots$$

The intersection of these spaces is the vector space of *infinitely* differentiable functions, denoted $C^\infty(\mathbb{R})$. Explicitly:

$$C^\infty(\mathbb{R}) = \bigcap_{k=0}^{\infty} C^k(\mathbb{R})$$

The vector space $C^\infty(\mathbb{R})$, though much smaller than $C(\mathbb{R})$, is still pretty big. It includes all polynomial functions (i.e., the image of $\mathbb{R}[x]$ in $F(\mathbb{R}, \mathbb{R})$ lives inside $C^\infty(\mathbb{R})$). It also includes rational functions where the denominator is never zero. It includes other functions involving exponentials, sines, and cosines, as long as these functions don't have zeros for their denominators.

5.2. **Differentiation as a linear transformation.** Differentiation can be defined as a linear transformation:

$$C^1(\mathbb{R}) \to C(\mathbb{R})$$

The following emerge from some thought:

- The *kernel* of this linear transformation is P_0, the space of constant functions. The kernel is one-dimensional.
- The linear transformation is surjective, i.e., its image is all of $C(\mathbb{R})$. This follows from the fact that every continuous function is the derivative of its integral.
- The fibers of differentiation, also called the "indefinite integral", are of the form (particular antiderivative) $+C$, where C is an arbitrary constant. The $+C$ arises from the fact that each fiber is a translate of the kernel, and the kernel is the space of constant functions.
- Note that $C(\mathbb{R})$ and $C^1(\mathbb{R})$ are infinite-dimensional spaces, and $C^1(\mathbb{R})$ is a proper subspace of $C(\mathbb{R})$. We thus have an interesting situation where there is a *surjective* linear transformation from a proper subspace to the whole space. Note that this kind of situation cannot arise with finite-dimensional spaces.
- For $k \geq 1$, the image of $C^k(\mathbb{R})$ under differentiation is $C^{k-1}(\mathbb{R})$. Moreover, the inverse image of $C^{k-1}(\mathbb{R})$ under differentiation is $C^k(\mathbb{R})$.
- The image of $C^\infty(\mathbb{R})$ under differentiation is $C^\infty(\mathbb{R})$. Moreover, the inverse image of $C^\infty(\mathbb{R})$ under differentiation is $C^\infty(\mathbb{R})$.
- The image of $\mathbb{R}[x]$ under differentiation is $\mathbb{R}[x]$. Moreover, the inverse image of $\mathbb{R}[x]$ under differentiation is also $\mathbb{R}[x]$. More detail: for $n \geq 1$, the image of P_n under differentiation is P_{n-1}, and the inverse image of P_{n-1} is P_n.
- The kernel of the k-fold iteration of differentiation is P_{k-1}, or rather, the image of P_{k-1} in $F(\mathbb{R}, \mathbb{R})$ (depending on whether we are thinking of differentiation as an operator $C^1(\mathbb{R}) \to C(\mathbb{R})$ or as an operator $\mathbb{R}[x] \to \mathbb{R}[x]$).
- We can also define a formal differentiation operator from $\mathbb{R}(x)$ to $\mathbb{R}(x)$ (recall the $\mathbb{R}(x)$ is not identifiable with a subspace of $C(\mathbb{R})$ because of the problem of denominator blow-up). The kernel is once again P_0. The image of $\mathbb{R}(x)$ under this differentiation operator is a subspace of $\mathbb{R}(x)$, but is *not* all of $\mathbb{R}(x)$. In other words, there do exist rational functions that do not have any rational function as their antiderivative. The function $1/(x^2 + 1)$, whose antiderivatives are all of the form $(\arctan x) + C$, is an example that is *not* in the image of $\mathbb{R}(x)$ under differentiation.

5.3. **Knowledge of derivatives and antiderivatives on a spanning set suffices.** Suppose V is a vector subspace of the vector space $C^\infty(\mathbb{R})$. We know that differentiation is linear. We now explore how that information is useful in computing the derivatives and antiderivatives of functions in V based on knowledge of derivatives and antiderivatives of functions in a spanning set S for V.

Let's tackle differentiation first. The first step is to express the function we want to differentiate as a linear combination of the functions in the spanning set S . Now, use the linearity of differentiation to express its derivative as the corresponding linear combination of the *derivatives* of functions in S.

For instance, suppose V is the span of sin and exp in $C^\infty(\mathbb{R})$ and we know that the derivative of sin is cos and the derivative of exp is exp. In this case, $S = \{\sin, \exp\}$. Then the derivative of the function:

$$f(x) = 2\sin x + 5\exp(x)$$

is:

$$f'(x) = 2\sin' x + 5\exp'(x) = 2\cos x + 5\exp(x)$$

More generally, given an arbitrary function:

$$f(x) = a\sin x + b\exp(x)$$

The derivative is:

$$f'(x) = a\cos x + b\exp(x)$$

Note that it is the fact of the functions *spanning* V that is crucial in allowing us to be able to write *any* function in V as a linear combination of the functions.

The computational procedure tells us a bit more: suppose S is a spanning set for a subspace V of $C^\infty(\mathbb{R})$. Suppose W is the image of V under differentiation. Then, the image of S under differentiation is a spanning set for W.

In particular, this means that if the image of S under differentiation lies inside V, then the image of V under differentiation is a subspace of V. This is what happens with the space $\mathbb{R}[x]$ of polynomials: the derivative of every power function (with nonnegative integer exponent) is a polynomial, and hence, the derivative of every polynomial is a polynomial.

Something similar applies to indefinite integration. We need to be a *little* more careful with indefinite integration, because the antiderivative is not unique. Instead, the indefinite integral of any function is a translate of the one-dimensional kernel. The obligatory $+C$ of indefinite integration is to account for the fact that the kernel of differentiation is the one-dimensional space of constant functions.

For instance, suppose we know that an antiderivative of sin is $-\cos$ and an antiderivative of exp is exp. Then, the indefinite integral of the function:

$$f(x) = 2\sin x + 5\exp(x)$$

is:

$$\int f(x)\,dx = 2(-\cos x) + 5\exp x + C$$

Let's consider some other settings where this idea has come in play:

- Consider $V = \mathbb{R}[x]$ and $S = \{1, x, x^2, \dots, \}$. The set S is a spanning set for V. In fact, S is a basis for V. We know how to diffferentiate any member of S: the power rule says that the derivative of x^n with respect to x is nx^{n-1}. This rule allows us to differentiate *any* polynomial, because a polynomial is a linear combination of these power functions. In other words, knowledge of how to differentiate functions in S tells us how to differentiate anything in V.
- Consder $V = \mathbb{R}[x]$ and $S = \{1, x, x^2, \dots, \}$ (same as above), but we are now interested in indefinite integration. Note that V already contains the constant functions (the kernel of differentiation) so we do not need to worry about the $+C$ taking us out of the space. The antiderivative of x^n is $x^{n+1}/(n+1)$. The formula is not correct at $n = -1$, but we are not considering that case here, since $n \geq 0$ here. We can use knowledge of antiderivatives of all functions in S to obtain antiderivatives of all functions in V. Further, since the antiderivatives of all elements of S are within V, every element of V has an antiderivative within V. And further, because constants are in V, *every* antiderivative of an element of V (aka a polynomial) is an element of V (aka a polynomial).
- Rational functions are somewhat similar: integration of rational functions relies on partial fractions theory, as discussed in the next section.

5.4. **How partial fractions help with integration.** Let's now revisit the topic of *partial fractions* as a tool for integrating rational functions. The idea behind partial fractions is to consider an integration problem with respect to a variable x with integrand of the following form:

$$\frac{a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}}{p(x)}$$

where p is a polynomial of degree n. For convenience, we may take p to be a monic polynomial, i.e., a polynomial with leading coefficient 1. For p fixed, the set of all rational functions of the form above forms a vector subspace of dimension n inside $\mathbb{R}(x)$. A natural choice of basis for this subspace is:

$$\frac{1}{p(x)}, \frac{x}{p(x)}, \ldots, \frac{x^{n-1}}{p(x)}$$

The goal of partial fraction theory is to provide an *alternate basis* for this space of functions with the property that those basis elements are particularly easy to integrate (recurring to one of our earlier questions). Let's illustrate one special case: the case that p has n distinct real roots $\alpha_1, \alpha_2, \ldots, \alpha_n$. The alternate basis in this case is:

$$\frac{1}{x - \alpha_1}, \frac{1}{x - \alpha_2}, \ldots, \frac{1}{x - \alpha_n}$$

The explicit goal is to rewrite a partial fraction:

$$\frac{a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}}{p(x)}$$

in terms of the basis above. If we denote the numerator as $r(x)$, we want to write:

$$\frac{r(x)}{p(x)} = \frac{c_1}{x - \alpha_1} + \frac{c_2}{x - \alpha_2} + \cdots + \frac{c_n}{x - \alpha_n}$$

The explicit formula is:

$$c_i = \frac{r(\alpha_i)}{\prod_{j \neq i} (\alpha_i - \alpha_j)}$$

Once we rewrite the original rational function as a linear combination of the new basis vectors, we can integrate it easily because we know the antiderivatives of each of the basis vectors. The antiderivative is thus:

$$\left(\sum_{i=1}^{n} \frac{r(\alpha_i)}{\prod_{j \neq i} (\alpha_i - \alpha_j)} \ln |x - \alpha_i| \right) + C$$

where the obligatory $+C$ is put for the usual reasons.

Note that this process only handles rational functions that are proper fractions, i.e., the degree of the numerator must be less than that of the denominator. For other rational functions, we first convert the "improper fraction" to a mixed fraction form using Euclidean division, then we integrate the polynomial part in the typical way that we integrate polynomials, and integrate the remaining proper fraction as above.

We now consider cases where p is a polynomial of a different type.

Suppose p is a monic polynomial of degree n that is a product of pairwise distinct irreducible factors that are all either monic linear or monic quadratic. Call the roots for the linear polynomials $\alpha_1, \alpha_2, \ldots, \alpha_s$ and call the monic quadratic factors $q_1, q_2, \ldots, q_t$. We are interested in finding an alternate basis of the space of all rational functions of the form $r(x)/p(x)$ where the degree of r is less than n.

This should be familiar to you from the halcyon days of doing partial fractions. For instance, consider the example where $p(x) = (x - 1)(x^2 + x + 1)$. In this case, the basis is:

$$\frac{1}{x - 1}, \frac{1}{x^2 + x + 1}, \frac{2x + 1}{x^2 + x + 1}$$

Note that an easy sanity check is that the *size* of the basis should be n. This is clear in the above example with $n = 3$, but let's reason generically.

We have that:

$$p(x) = \left[\prod_{i=1}^{s}(x - \alpha_i)\right]\left[\prod_{j=1}^{t} q_j(x)\right]$$

By degree considerations, we get that:

$$s + 2t = n$$

Now, the vector space for which we are trying to obtain a basis has dimension n. This means that the basis we are looking for should have size n, as it does, because we have a basis vector $1/(x - \alpha_i)$ for each linear factor $x - \alpha_i$ (total of s basis vectors here) and we have two basis vectors $1/q_j(x)$ and $q_j'(x)/q_j(x)$ for each quadratic factor $q_j(x)$ (total of $2t$ basis vectors here).

Now, recall that the reciprocals of the linear factors integrate to logarithms. The expressions of the form $1/q_j(x)$ integrate to an expression involving arctan. The expressions of the form $q_j'(x)/q_j(x)$ integrate to logarithms.

6. Spaces of sequences

6.1. Spaces of sequences of real numbers. Recall that a sequence of real numbers is a function:

$$f : \mathbb{N} \to \mathbb{R}$$

We typically denote the function input as a subscript, so for instance we may denote $f(n)$ by a_n. The sequence is typically written by listing its elements, so the above function is written as:

$$f(1), f(2), f(3), \dots$$

or equivalently as:

$$a_1, a_2, a_3, \dots$$

We've already discussed that the space of all functions from *any* set to the reals has a natural structure of a vector space. The addition is pointwise: we add the values of the functions at each point. The scalar multiplication is also pointwise. For sequences, let's describe these operations more explicitly:

- The zero sequence is the sequence all of whose entries are zero.
- Given two sequences $a_1, a_2, a_3, \dots$ and $b_1, b_2, b_3, \dots$, the sum of the sequences is the sequence $a_1 + b_1, a_2 + b_2, a_3 + b_3, \dots$.
- Given a sequence $a_1, a_2, a_3, \dots$ and a real number λ, the sequence we get after scalar multiplication is $\lambda a_1, \lambda a_2, \lambda a_3, \dots$.

The vector space of sequences is infinite-dimensional.

6.2. Formal power series and convergence. A *formal power series* is defined as a series of the form:

$$\sum_{i=0}^{\infty} a_i x^i$$

where x is an indeterminate, and a_i are all real numbers. In other words, formal power series are like polynomials in x, except that they *can* go on indefinitely rather than being forced to terminate at a finite stage. Examples are:

$$1 + x + x^2 + \dots$$

$$1 - x + x^2 - x^3 + x^4 - \dots$$

The set of all formal power series forms a vector space under coefficient-wise addition and scalar multiplication. Note that this vector space looks a lot like (in fact, is isomorphic to) the vector space of sequences. The only difference is that we write the "vectors" as formal sums rather than as comma-separated lists.

The vector space of all formal power series is denoted $\mathbb{R}[[x]]$. Note that this vector space has a lot of additional structure to it beyond simply being a vector space.

Several statements about the nature of Taylor series and power series summation operators, which you have encountered in the past, can be framed as statements about the injectivity, surjectivity, kernel, and image of suitably defined linear transformations.

6.3. **Differentiation as an operator from formal power series to itself.** We can define a formal differentiation operator:

$$\mathbb{R}[[x]] \to \mathbb{R}[[x]]$$

that is done term-wise. Explicitly, the constant term falls off, and each $a_n x^n$ for $n \geq 1$ gets differentiated to $n a_n x^{n-1}$. In other words, the new coefficient for x^{n-1} is n times the old coefficient for x^n. Thus:

$$\frac{d}{dx} \sum_{i=0}^{\infty} a_i x^i = \sum_{i=0}^{\infty} (i+1) a_{i+1} x^i$$

Formal differentiation is a linear transformation. The kernel is the space of constant power series. These are power series where all coefficients $a_i, i \geq 1$ are equal to 0. The coefficient a_0 may be zero or nonzero. This kernel is one-dimensional, so formal differentiation is not injective.

On the other hand, formal differentiation is *surjective*. Every formal power series arises as the formal derivative of a formal power series. The "indefinite integral" is non-unique (we have flexibility in the choice of constant term, with the famed $+C$ out there). But it can be computed by term-wise integration. Explicitly:

$$\int \left(\sum_{i=0}^{\infty} a_i x^i \right) dx = \left(\sum_{i=1}^{\infty} \frac{a_{i-1}}{i} x^i \right) + C$$

6.4. **Taylor series operator and power series summation operator.** The *Taylor series* operator is an operator whose domain is a space of function-like things and whose co-domain is a space of power series-like things. For our purposes, we choose a simple implementation, albeit not the best one. We view the Taylor series operator as an operator:

$$\text{Taylor series} : C^{\infty}(\mathbb{R}) \to \mathbb{R}[[x]]$$

The operator takes as input an infinitely differentiable function defined on all of $\mathbb{R}$ and outputs the Taylor series of the function centered at 0. Explicitly, the operator works as follows:

$$f \mapsto \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$$

Note that we need infinite differentiability at 0 in order to make sense of the expression. We do not really need infinite differentiability on all of $\mathbb{R}$, so the domain is in some sense too restrictive. But let's not get into the issue of trying to define new spaces to overcome this space.

For the power series summation operator, let Ω be the subspace of $\mathbb{R}[[x]]$ comprising the power series that converge globally. Then, we have a mapping:

$$\text{Summation} : \Omega \to C^{\infty}(\mathbb{R})$$

that sends a formal power series to the function to which it converges. It turns out to be true that if we start with an element of Ω, apply the summation operator to it, and then take the Taylor series of that, we land up with the same thing we started with. On the other hand, if we start with something in $C^{\infty}(\mathbb{R})$, take its Taylor series, and then try summing that up, we may land up with a completely different function, if we end up with a function at all. You should remember more of this from your study of Taylor series and power series in single variable calculus.

7. Differential operators and differential equations

7.1. **Linear differential operators with constant coefficients.** Denote by D the operator of differentiation. Note that D can be viewed as an operator in many contexts:

- It defines a linear transformation $C^1(\mathbb{R}) \to C(\mathbb{R})$
- It defines a linear transformation $C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$
- It defines a linear transformation $\mathbb{R}[x] \to \mathbb{R}[x]$
- It defines a linear transformation $\mathbb{R}(x) \to \mathbb{R}(x)$
- It defines a linear transformation $\mathbb{R}[[x]] \to \mathbb{R}[[x]]$

We will consider the case of $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$. We will use I to denote the identity transformation $C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$. We can then construct other linear transformations, loosely called *linear differential operators with constant coefficients*, by adding, multiplying, and scalar multiplying these. In other words, we can use polynomials in D. For instance:

- $I + D$ is the linear transformation $f \mapsto f + f'$.
- $I - D$ is the linear transformation $f \mapsto f - f'$.
- $I + 2D + 3D^2$ is the linear transformation $f \mapsto f + 2f' + 3f''$.

Finding the *kernel* of any such differential operator is equivalent to solving a homogeneous linear differential equation with constant coefficients. Finding the *fiber* over a specific nonzero function in $C^\infty(\mathbb{R})$ under any such differential operator is equivalent to solving a non-homogeneous linear differential equation with constant coefficients.

For instance, finding the kernel of $I + D$ is equivalent to solving the following linear differential equation, where x is the independent variable and y is the dependent variable. The linear differential equation is:

$$y + y' = 0$$

We know that this solves to:

$$y = Ce^{-x}, C \in \mathbb{R}$$

The kernel is thus the one-dimensional space spanned by the function e^{-x}.

On the other hand, if we are trying to find the fiber of the function, say $x \mapsto x^2$, that is equivalent to solving the non-homogeneous linear differential equation:

$$y + y' = x^2$$

A *particular solution* here is $y = x^2 - 2x + 2$. The general solution is the translate of the kernel by the particular solution, i.e., the general solution function is:

$$x^2 - 2x + 2 + Ce^{-x}, C \in \mathbb{R}$$

What this means is that each specific value of C gives a different particular solution. Pictorially, each such solution function is a point and the fiber we are thinking of is the line of all such points.

7.2. **The first-order linear differential equation: a full understanding.** We now move to the somewhat more general setting of first-order linear differential operators with nonconstant coefficients.

Consider a first-order linear differential equation with independent variable x and dependent variable y, with the equation having the form:

$$y' + p(x)y = q(x)$$

where $p, q \in C^\infty(\mathbb{R})$.

We solve this equation as follows. Let H be an antiderivative of p, so that $H'(x) = p(x)$.

$$\frac{d}{dx}\left(ye^{H(x)}\right) = q(x)e^{H(x)}$$

This gives:

$$ye^{H(x)} = \int q(x)e^{H(x)}\,dx$$

So:

$$y = e^{-H(x)} \int q(x)e^{H(x)}\,dx$$

The indefinite integration gives a $+C$, so overall, we get:

$$y = Ce^{-H(x)} + \text{particular solution}$$

It's now time to understand this in terms of linear algebra.

Define a linear transformation $L : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ as:

$$f(x) \mapsto f'(x) + p(x)f(x)$$

This is a differential operator of a more complicated sort than seen earlier (explicitly, it is $L = D + pI$). The kernel of L is the solution set when $q(x) = 0$, and this just becomes the set:

$$Ce^{-H(x)}, C \in \mathbb{R}$$

In other words, each value of C gives a solution, i.e., a "point" in the kernel. The entire solution set is a line, i.e., a one-dimensional space, spanned by the function $e^{-H(x)}$.

The fiber over a function q is a translate of this kernel:

$$y = Ce^{-H(x)} + \text{particular solution}$$

Note that the fiber is non-empty because a particular solution can be obtained by integration. Thus, the linear transformation L is surjective. L is not injective because it has a one-dimensional kernel.

7.3. **Higher order differential operators and differential equations.** Here's a brief description of the theory of differential operators and differential equations that is relevant here.

Consider a linear differential equation of order n of the following form:

$$y^{(n)} + p_{n-1}(x)y^{(n-1)} + \cdots + p_1(x)y' + p_0(x)y = q(x)$$

The left side can be viewed as a linear differential operator of order n from $C^\infty(\mathbb{R})$ to $C^\infty(\mathbb{R})$. Explicitly, it is the operator:

$$L(y) = y^{(n)} + p_{n-1}(x)y^{(n-1)} + \cdots + p_1(x)y' + p_0(x)y$$

Here, all the functions $p_0, p_1, \dots, p_{n-1}, q$ are in $C^\infty(\mathbb{R})$.

The kernel of this operator is the solution set to the corresponding homogeneous linear differential equation:

$$y^{(n)} + p_{n-1}(x)y^{(n-1)} + \cdots + p_1(x)y' + p_0(x)y = 0$$

It turns out that we generically expect this kernel to be a n-dimensional subspace of $C^\infty(\mathbb{R})$. Explicitly, this means that the kernel has a basis comprising n functions, each of which is a solution to the system. This is consistent with the general principle that we expect n independent parameters in the general solution to a differential equation of order n.

The general solution to the original non-homogeneous linear differential equation, if it exists, is of the form:

(Particular solution) + (Arbitrary element of the kernel)

The elements of the kernel are termed "auxilliary solutions" so we can rewrite this as:

(General solution) = (Particular solution) + ((General) auxilliary solution)

The parameters (freely floating constants) all come from the choice of arbitrary element of the kernel. There are n of them, as expected. Unfortunately, unlike the first-order case, there is no generic integration formula for finding the particular solution. The first-order and second-order cases are the only cases where a generic integration formula is known for finding the auxilliary solutions. Also, it is known how to find the auxilliary solutions in the constant coefficients case for arbitrary order.

The linear differential operator L is thus *surjective but not injective*: it has n-dimensional kernel.