10 Linear transformations

Together with the vector spaces, the second most important notion of linear algebra is the notion of a *linear transformation*, which is a map from one vector space to another satisfying certain "linear" conditions. Before introducing formally linear transformations, I consider a very general notion of a map.

10.1 Functions, or maps, from a set to a set

Let X, Y be two sets. By definition, a *map*, or a *function*, f from X to Y, which is usually denoted as

$$f\colon X \longrightarrow Y,$$

is a rule that to each element $x \in X$ assigns an element $y \in Y$, which is quite often written as $y = f(x)$. Note that the definition of a function involves three elements: the rule itself, the set X, which is called *domain* of f, and the set Y. It is possible to have *different* functions with the same rule but different, e.g., domains. The *range* of f is the set of all such y from Y such that there is $x \in X$ for which $f(x) = y$. Note that the range of a function does not have to coincide with Y and can be a proper subset of Y.

If $f\colon X \longrightarrow Y$ and $g\colon Y \longrightarrow Z$ then it is possible to define function *composition*, which is a map from X to Z, and which is denoted $g \circ f$:

$$g \circ f\colon X \longrightarrow Z,$$

by specifying that $(g \circ f)(x) = g(f(x))$.

Proposition 10.1. *Function composition if associative. That is, if $f\colon X \longrightarrow Y$, $g\colon Y \longrightarrow Z$, and $h\colon Z \longrightarrow W$ then*

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

Exercise 1. Prove the proposition by evaluating the expressions on the left and the right at an arbitrary $x \in X$.

A function $f\colon X \longrightarrow Y$ is called an *injection*, or *one-to-one*, if $f(x_1) = f(x_2)$ implies $x_1 = x_2$, a *surjection*, or *onto*, if for any $y \in Y$ there is $x \in X$ such that $f(x) = y$ (that is, the range of f is equal to Y), and a *bijection* if it is both one-to-one and onto (sometimes it is called *one-to-one correspondence*). To formulate the following result I introduce the identity map $1_X\colon X \longrightarrow X$, which is defined by the rule $1_X(x) = x$.

Theorem 10.2. *Let $f\colon X \longrightarrow Y$, where X, Y are nonempty sets. f is injective if and only if there exists a map $g\colon Y \longrightarrow X$ such that $g \circ f = 1_X$. f is surjective if and only if there exists a map $g\colon Y \longrightarrow X$ such that $f \circ g = 1_Y$, and f is bijective if and only if there exists a map $g\colon Y \longrightarrow X$ such that $f \circ g = 1_Y$ and $g \circ f = 1_X$. (In the case of the bijection f function g is usually called the inverse of f and denoted f^{-1}.)*

Math 329: *Intermediate Linear Algebra* by Artem Novozhilov©
e-mail: artem.novozhilov@ndsu.edu. Fall 2016

Remark 10.3. Even if the reader never heard about injections, surjections, and bijections, she should at least at some intuitive level see the resemblance to the notions of left and right inverses, and simply the inverse.

Proof. Assume that f is injective. Now I will construct my g. If $y \in Y$ is such that there is $x \in X$ for which $f(x) = y$ I put $g(y) = x$, which can be done uniquely due to injectivity of f. For those y for which there is no such x, I pick $x_0 \in X$ since X is nonempty, and put $g(y) = x_0$. Now take any $x \in X$, by construction $g(f(x)) = x$ as expected. In the other direction, assume that $g \circ f = 1_X$ and take $x_1, x_2 \in X$ such that $f(x_1) = f(x_2)$, applying g I get $x_1 = x_2$ and hence f is injective.

Assume f is surjective. This means that for any $y \in Y$ there are some $x \in X$ for which $f(x) = y$. Pick any of these x and define $g(y) = x$. By construction $f(g(y)) = y$ as expected. Conversely, if $f \circ g = 1_Y$ then $f(g(y)) = y$ for any y, and hence for any $y \in Y$ there is $x \in X$ for which $f(x) = y$, hence f is surjective.

The last part of the theorem follows from putting together the previous two parts and proving that these two g in this case coincide. ■

Exercise 2. Come up with examples of real values functions (that is, with the functions with which you mostly dealt with in calculus) which is 1) injection but not surjection, 2) surjection but not injection, 3) bijection.

Exercise 3. If the function composition commutative? I.e., is $f \circ g = g \circ f$ for all f, g?

10.2 Linear transformations. The dimension formula

Now I am ready to define a linear transformation $\mathscr{A} \colon U \longrightarrow V$.

Definition 10.4. *A linear transformation $\mathscr{A}$ from vector space U over $\mathbf{F}$ to vector space V over the same field $\mathbf{F}$ is a map*

$$\mathscr{A} \colon U \longrightarrow V,$$

that satisfies the following two properties:

$$\mathscr{A}(u + v) = \mathscr{A}(u) + \mathscr{A}(v), \quad \text{for all } u, v \in U,$$
$$\mathscr{A}(\alpha u) = \alpha \mathscr{A}(u), \quad \text{for all } u \in U,\ \alpha \in \mathbf{F}.$$

Note that for linear transformation $\mathscr{A}$ it immediately follows (e.g., by induction) that

$$\mathscr{A}(\alpha_1 v_1 + \ldots + \alpha_k v_k) = \alpha_1 \mathscr{A}(v_1) + \ldots + \alpha_k \mathscr{A}(v_k),$$

that is linear transformations map linear combinations to linear combinations.

Example 10.5. Let $U = V = \mathbf{R}$. Then the only linear transformation $\mathscr{A}$ is given by a multiplication by a constant:

$$\mathscr{A}(x) = \alpha x.$$

It is trivial to check that if $\mathscr{A}(x) = \alpha x$ then $\mathscr{A}$ is linear. Conversely, represent any element of $\mathbf{R}$ as a linear combination of its basis, say, (1). I have $x = x \cdot 1$, and using the properties of the linear transformation $\mathscr{A}(x \cdot 1) = x\mathscr{A}(1) = \alpha x$ since $\mathscr{A}(1)$ is a scalar.

Note that usually defined as being *linear* function $f(x) = ax + b$ is not a linear transformation by the definition above.

Example 10.6. Let $U = \mathbf{R}^n$, $V = \mathbf{R}^m$ and $\mathscr{A} = \boldsymbol{A} = [a_{ij}]_{m\times n}$. Then multiplication by $\boldsymbol{A}$ is a linear transformation, since

$$\boldsymbol{A}(\alpha \boldsymbol{v} + \boldsymbol{u}) = \alpha \boldsymbol{A}\boldsymbol{v} + \boldsymbol{A}\boldsymbol{u}$$

by the properties of matrix multiplication.

Example 10.7. Let $U = \mathbf{P}_n, V = \mathbf{P}_{n-1}$, where $\mathbf{P}_n$ is the real vector space of polynomials of degree at most n. Let $\mathscr{A} = \frac{\mathrm{d}}{\mathrm{d}x}$, i.e., differentiation. Then $\mathscr{A}$ is a linear transformation, since for any $p, q \in \mathbf{P}_n$

$$\mathscr{A}(\alpha p + q) = \alpha \mathscr{A}(p) + \mathscr{A}(q)$$

by the properties of differentiation.

Example 10.8. Let $U = V = \mathbf{R}^2$ and let $\mathscr{A}$ be a rotation of every vector in the plane counterclockwise by an angle θ. Make a graph to convince yourself that this is indeed linear transformation.

Here are some simple properties of linear transformations:

- If $\mathscr{A}: U \longrightarrow V$ is a linear transformation then $\mathscr{A}(0) = 0$ (note that the zeros are from different vector spaces). Indeed $\mathscr{A}(0) = \mathscr{A}(0+0) = \mathscr{A}(0) + \mathscr{A}(0) \implies \mathscr{A}(0) = 0$.
- Let $\mathscr{A}: U \longrightarrow V, \mathscr{B}: V \longrightarrow W$ be linear transformations on the vector spaces over the same field. Then $\mathscr{B}\circ\mathscr{A}: U \longrightarrow W$ is a linear transformation. Indeed, $(\mathscr{B}\circ\mathscr{A})(\alpha u + v) = \mathscr{B}(\mathscr{A}(\alpha u + v)) = \mathscr{B}(\alpha\mathscr{A}(u) + \mathscr{A}(v)) = \alpha\mathscr{B}(\mathscr{A}(u)) + \mathscr{B}(\mathscr{A}(v)) = \alpha(\mathscr{B}\circ\mathscr{A})(u) + (\mathscr{B}\circ\mathscr{A})(v)$.
- Let $\mathscr{A}: U \longrightarrow V$ be a linear transformation. Then it is injective if and only if $\mathscr{A}(u) = 0$ implies that $u = 0$. Indeed, if $\mathscr{A}$ is injective and the fact that $\mathscr{A}(0) = 0$ the conclusion follows. Now, assume that $\mathscr{A}(u) = 0 \implies u = 0$. Consider $\mathscr{A}(u_1) = \mathscr{A}(u_2)$. By linearity $\mathscr{A}(u_1 - u_2) = 0$, by assumption $u_1 - u_2 = 0$, hence $u_1 = u_2$ and hence $\mathscr{A}$ is injective.
- Let $\mathscr{A}: U \longrightarrow V$ be a linear transformation, which is bijective. That is there is the inverse $g: V \longrightarrow U$ such that $g(\mathscr{A}(u)) = u \in U$ and $\mathscr{A}(g(v)) = v \in V$. Then g is a linear transformation. Indeed, to show that $g(\alpha v_1 + v_2) = \alpha g(v_1) + g(v_2)$ I will use the injectivity of $\mathscr{A}$ to show an equivalent equality $\mathscr{A}(g(\alpha v_1 + v_2)) = \mathscr{A}(\alpha g(v_1) + g(v_2))$. The latter equality holds by the linearity of $\mathscr{A}$.

Finally, I can rigorously define the isomorphism, which I already used in the previous lecture.

Definition 10.9. *A bijective linear transformation $\mathscr{A}: U \longrightarrow V$ is called an isomorphism. Two vector spaces for which there is an isomorphism are called isomorphic.*

Here are several useful statements, using the notion of an isomorphism, whose proofs are left as exercises.

Let $\mathscr{A}: U \longrightarrow V$ be a linear transformation between finite dimensional vector spaces over $\mathbf{F}$. Then

- Let $\mathscr{A}$ be a linear transformation. Then it is an isomorphism if and only if the equation $\mathscr{A}(u) = v$ has a unique solution for all $v \in V$.
- Let $\mathscr{A}$ be an isomorphism. Then $(u_1, \ldots, u_n)$ is a basis of U if and only if $(\mathscr{A}(u_1), \ldots, \mathscr{A}(u_n))$ is a basis of V.

- Let $(u_1, \dots, u_n)$ be a basis of U and $(v_1, \dots, v_n)$ be a basis of V. If $v_j = \mathscr{A}(u_j)$ then $\mathscr{A}$ is an isomorphism.

In particular, the last point shows that any finite dimensional vector spaces over the same field $\mathbf{F}$ of the same dimension are isomorphic, and, in particular, isomorphic to $\mathbf{F}^n$.

With any linear transformation two subsets are identified: *kernel* of $\mathscr{A}$ and *image* of $\mathscr{A}$ (they are also called *null space* and *range* of $\mathscr{A}$).

Definition 10.10. *Let $\mathscr{A}\colon U \longrightarrow V$ be a linear transformation. Then its kernel, $\ker \mathscr{A}$, and image, $\operatorname{im} \mathscr{A}$, are defined as follows:*

$$\ker \mathscr{A} = \{u \in U\colon \mathscr{A}(u) = 0\},$$
$$\operatorname{im} \mathscr{A} = \{v \in V\colon \text{ there is } u \in U \text{ such that } \mathscr{A}(u) = v\}.$$

Proposition 10.11. $\ker \mathscr{A}$ *and* $\operatorname{im} \mathscr{A}$ *are subspaces of U and V respectively.*

Exercise 4. Prove this proposition.

Proposition 10.12. *Let $\mathscr{A}\colon U \longrightarrow V$ be a linear transformation. Then it is surjective if and only if $\operatorname{im} \mathscr{A} = V$.*

Exercise 5. Prove the last proposition.

The dimension of the image of $\mathscr{A}$ is called the *rank* of linear transformation, and the dimension of the kernel of $\mathscr{A}$ is called the *nullity* of $\mathscr{A}$.

Next is one of the most important results of linear algebra, which we actually already saw in a matrix disguise.

Theorem 10.13. *Let $\mathscr{A}\colon U \longrightarrow V$ be a linear transformation between finite dimensional vector spaces. Then*

$$\dim \ker \mathscr{A} + \dim \operatorname{im} \mathscr{A} = \dim U.$$

Proof. Assume that $\dim U = n$. Since $\ker \mathscr{A} \subseteq U$ then $\dim \ker \mathscr{A} = k \leq n$. Let $(u_1, \dots, u_k)$ be a basis of $\ker \mathscr{A}$. I extend it to the basis of U by adding more vectors $(u_1, \dots, u_k, v_1, \dots, v_{n-k})$. For $j = 1, \dots, n-k$ let $w_j = \mathscr{A}(v_j)$. I claim that $S = (w_1, \dots, w_{n-k})$ is a basis of $\operatorname{im} \mathscr{A}$. To show this I must show that S spans the image and that it is a linearly independent collection.

First, let $w \in \operatorname{im} \mathscr{A}$, that is that is $u \in U$ such that $w = \mathscr{A}(u)$. Since $(u_1, \dots, u_k, v_1, \dots, v_{n-k})$ is a basis, $u = \alpha_1 u_1 + \ldots + \alpha_k u_+ \beta_1 v_1 + \ldots + \beta_{n-k} v_k$. Using the properties of $\mathscr{A}$ I get $w = \beta_1 \mathscr{A}(v_1) + \ldots \beta_{n-k} \mathscr{A}(v_{n-k}) = \beta_1 w_1 + \ldots + \beta_{n-k} w_{n-k}$, which proves that S spans the image.

Now take the linear combination

$$\gamma_1 w_1 + \ldots + \gamma_{n-k} w_{n-k} = 0.$$

Take $v = \gamma_1 v_1 + \ldots + \gamma_{n-k} v_{n-k}$, where v_j are the vectors from the basis of U. This construction implies that $\mathscr{A}(v) = \gamma_1 w_1 + \ldots + \gamma_{n-k} w_{n-k} = 0$, i.e., $v \in \ker \mathscr{A}$ and hence $v = \alpha_1 u_1 + \ldots + \alpha_k u_k$. Then, finally, I have that

$$-v + v = -\alpha_1 u_1 - \ldots - \alpha_k u_k + \gamma_1 v_1 + \ldots + \gamma_{n-k} v_{n-k} = 0,$$

and since $(u_1, \dots, u_k, v_1, \dots, v_{n-k})$ is a basis all the scalars must be zero, including $\gamma_1 = \ldots = \gamma_{n-k} = 0$, hence S is linearly independent.

Now counting the number of vectors in each basis finishes the proof. ∎

Remark 10.14. Since matrices are examples of linear transformations, all the information we said about linear transformations if true for them. In particular, we now can immediately conclude that for a matrix to be invertible (to be an isomorphisms) we must have that the number of rows is equal to the number of columns, that it, the matrix must be square. Indeed, since for the invertible matrix $\ker \boldsymbol{A} = \{0\}$ and $\operatorname{rank} \boldsymbol{A} = \dim V$ by the dimension formula, therefore $\dim U = \dim V$ and hence $\boldsymbol{A}$ is square.

Example 10.15. Consider a linear transformation $\boldsymbol{A}\colon \mathbf{R}^5 \longrightarrow \mathbf{R}^4$, which is matrix multiplication with the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 & 1 \\ 3 & 3 & 3 & 3 & 2 \\ 1 & 1 & -1 & -1 & 0 \end{bmatrix}.$$

Let me find the image of $\boldsymbol{A}$, its kernel, its rank and nullity. First, let me explain what I mean when I ask to find, say, an image. We now know that an image is a subspace of $\mathbf{R}^4$ and therefore to find it means to produce a basis of it. The same about the kernel. It turns out that everything can be inferred from the reduced row echelon form of $\boldsymbol{A}$. Indeed, consider

$$\begin{bmatrix} 1 & 1 & 2 & 2 & 1 \\ 2 & 2 & 1 & 1 & 1 \\ 3 & 3 & 3 & 3 & 2 \\ 1 & 1 & -1 & -1 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 0 & 0 & 1/3 \\ 0 & 0 & 1 & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Therefore, I immediately can write down the general solution to $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$ as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = s \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + q \begin{bmatrix} -1/3 \\ 0 \\ -1/3 \\ 0 \\ 1 \end{bmatrix}.$$

Clearly three vectors on the right span $\ker \boldsymbol{A}$, but they are also linearly independent (look just at the entries corresponding to the free variables x_2, x_4, x_5, they form an identity matrix), and therefore form a basis of $\ker \boldsymbol{A}$. Immediately, $\dim \ker \boldsymbol{A} = 3$, which is the nullity of my linear transformation.

From the dimension formula I have that $\dim \operatorname{im} \boldsymbol{A} = 2$, which is the rank of my linear transformation (and this is just the number of pivots in the reduced row echelon form), but how to find a basis for it? Of course, I could transpose my matrix, row reduce it and take nonzero rows as the sought basis, however, I do not need to do this! I claim that if I take columns 1 and 3 (these are the columns with the pivots in the row reduced form) of the *original* matrix $\boldsymbol{A}$, then they form a basis of $\operatorname{im} \boldsymbol{A}$. That is

$$\operatorname{span} \operatorname{im} \boldsymbol{A} = \operatorname{span}\{[1\ 2\ 3\ 1], [2\ 1\ 3\ -1]\}.$$

How can I prove the last statement? First, convince yourself that these columns in the row reduced echelon form are indeed the basis for the column space: they are linearly independent and any other column can be represented as a linear combination of them. Second, recall that the row reduction amounts to multiplication by an invertible matrix from the left, and now we know that application of an invertible transformation (i.e., of an isomorphism) does not change linear dependence and/or independence. I invite the student to write down a formal argument along these lines.

10.3 The general solution of the linear equation

Let $\mathscr{A}\colon U \longrightarrow V$ be a linear operator. The equation

$$\mathscr{A}(x) = b$$

is called linear non-homogeneous if $b \neq 0$ and linear homogeneous if $b = 0$.

Remark 10.16. Please note that the linear equation that I consider is a more general object than the systems of linear algebraic equations I studied before. For example, differential equation

$$y' = y + \sin x$$

can be written in the form

$$\mathscr{A}(y) = f,$$

if I set $\mathscr{A} = \frac{\mathrm{d}}{\mathrm{d}x} - 1$ and $f = \sin$.

Proposition 10.17. *Consider the linear equation $\mathscr{A}(x) = b$. Assume that vector x_n solves it, and let H be the set of all solution of the homogeneous equation $\mathscr{A}(x) = 0$. Then all the solution to $\mathscr{A}(x) = b$ are given by*

$$x_g = x_n + x_h, \quad h_h \in H.$$

Proof. First, due to linearity of $\mathscr{A}$ it is clear that any vector of the form $x_n + x_h$ solves the equation. Now assume that x_g is an arbitrary solution, consider $x_g - x_n$, which, by construction, is in H, and therefore $x_g = x_n + x_h$. ■

Remark 10.18. This is a very elementary and yet important theorem. In particular, if $\ker \mathscr{A}$ is finite dimensional, all we need to do is to 1) find a basis of $\ker \mathscr{A}$ and 2) find any solution to the nonhomogeneous equation.

10.4 Matrix of a linear transformation

Now let me concentrate for a second on the linear transformations from $\mathbf{R}^n$ to $\mathbf{R}^m$ (or, more generally, from $\mathbf{F}^n$ to $\mathbf{F}^m$). I claim that *any* linear transformation $\mathscr{A}\colon \mathbf{R}^n \longrightarrow \mathbf{R}^m$ can be *represented* as multiplication by an $m \times n$ matrix $\boldsymbol{A}$. Indeed, let me take the standard basis of $\mathbf{R}^n$: $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$. Any vector $\boldsymbol{x} \in \mathbf{R}^n = x_1\boldsymbol{e}_1 + \ldots + x_n\boldsymbol{e}_n$. Now, consider a linear transformation $\mathscr{A}$ applied to $\boldsymbol{x}$. I get

$$\mathscr{A}(\boldsymbol{x}) = \mathscr{A}(x_1\boldsymbol{e}_1 + \ldots + x_n\boldsymbol{e}_n) = x_1\mathscr{A}(\boldsymbol{e}_1) + \ldots + x_n\mathscr{A}(\boldsymbol{e}_n) = x_1\boldsymbol{v}_1 + \ldots + x_n\boldsymbol{v}_n.$$

Therefore, if I join the vectors $\boldsymbol{v}_j$ together into the matrix $\boldsymbol{A} = [\boldsymbol{v}_1 \mid \ldots \mid \boldsymbol{v}_n]$ then the action of my linear transformation is exactly the multiplication by $\boldsymbol{A}$. And my matrix $\boldsymbol{A}$ is the matrix of my linear transformation in the standard basis. (Note that I could have gone in a different way, *defining* the multiplication of a vector by a matrix to satisfy the expression above.)

Example 10.19. What is the matrix of the linear transformation (operator) in Example 10.8? To see it we must determine the action of this transformation on the standard basis. In this case I get, by elementary trigonometric formulas, that

$$\mathscr{A}(\boldsymbol{e}_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \quad \mathscr{A}(\boldsymbol{e}_2) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix},$$

hence the matrix of the transformation $\mathscr{A}$ is

$$\boldsymbol{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Example 10.20. Consider the linear transformation $\mathscr{R}$ of $\mathbf{R}^2$ to $\mathbf{R}^2$, which is a reflection around x_1-axis. Of course, I need to prove that this transformation is linear, but it is much easier to see that the formula for this transformation is

$$\mathscr{R}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix},$$

which clearly satisfies the axioms of the linear transformation. The matrix of this transformation is

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Now consider the following question: Let $\mathscr{A}\colon \mathbf{R}^n \longrightarrow \mathbf{R}^m$ and $\mathscr{B}\colon \mathbf{R}^m \longrightarrow \mathbf{R}^l$ be linear transformations. According to the theory above these transformations can be represented by matrices $\boldsymbol{A}_{m\times n}$ and $\boldsymbol{B}_{l\times m}$. What is the matrix $\boldsymbol{C}$ of the composition $\mathscr{B}\circ\mathscr{A} = \mathscr{C}$? Clearly, I can find $\boldsymbol{C}$ by considering how $\mathscr{C}$ acts on the standard vectors:

$$\boldsymbol{C} = [\mathscr{C}(\boldsymbol{e}_1) \mid \ldots \mid \mathscr{C}(\boldsymbol{e}_n)] = [\mathscr{B}(\mathscr{A}(\boldsymbol{e}_1)) \mid \ldots \mid \mathscr{B}(\mathscr{A}(\boldsymbol{e}_n))].$$

Now let me consider only the first column. I know that action of $\mathscr{A}$ amounts to multiplication by $\boldsymbol{A}$, and similar for $\mathscr{B}$, hence

$$\boldsymbol{B}(\boldsymbol{A}\boldsymbol{e}_1) = \begin{bmatrix} b_{11} & \ldots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{l1} & \ldots & b_{lm} \end{bmatrix} \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} = \begin{bmatrix} b_{11}a_{11} + b_{12}a_{21} + \ldots + b_{1m}a_{m1} \\ \vdots \\ b_{l1}a_{11} + b_{l2}a_{21} + \ldots + b_{lm}a_{m1} \end{bmatrix},$$

which is the first column of matrix $\boldsymbol{C}$. Do you recognize the patten? Now I can conclude that the matrix of the composition of transformations is actually the product of the corresponding matrices! Even more importantly, the matrix multiplication can be *defined* so that the matrix of the composition of linear transformations be equal to the product of the corresponding matrices.

As a small, but pleasant, surprise we get free of charge

Proposition 10.21. *Matrix multiplication is associative.*

Now consider a linear transformation between abstract vector spaces $\mathscr{A}\colon U \longrightarrow V$ and assume that $\dim U = n$ and $\dim V = m$. Using the fact that these vector spaces are isomorphic to $\mathbf{F}^n$ and $\mathbf{F}^m$ respectively, I can expect that, by fixing the bases in U (denote $\mathcal{U}$) and V (denote $\mathcal{V}$), I can always represent my linear transformation as a matrix $\boldsymbol{A}_{m\times n} = [\mathscr{A}]_{\mathcal{U}}^{\mathcal{V}}$, note the dependence on the both bases.

Indeed, the following theorem holds.

Theorem 10.22. *Let $\mathscr{A}\colon U \longrightarrow V$ be a linear transformation, and let $\mathcal{U} = (u_1, \ldots, u_n)$ be a basis of U and $\mathcal{V} = (v_1, \ldots, v_m)$ be a basis of V. Then for any vector $u \in U$ and its image $v = \mathscr{A}(u) \in V$ there is an $m \times n$ matrix $\boldsymbol{A}$ with the property*

$$[v]_{\mathcal{V}} = \boldsymbol{A}[u]_{\mathcal{U}}.$$

Matrix $\boldsymbol{A}$ is called the matrix of linear transformation $\mathscr{A}$ with respect to the bases $\mathcal{U}$ and $\mathcal{V}$.

Remark 10.23. The statement of the theorem can be written in the following symmetric form

$$[\mathscr{A}(u)]_{\mathcal{V}} = [\mathscr{A}]_{\mathcal{U}}^{\mathcal{V}}[u]_{\mathcal{U}}.$$

Proof. Consider

$$\mathscr{A}(u_j) = a_{1j}v_1 + \ldots + a_{mj}v_j.$$

Form the matrix $\boldsymbol{A} = [a_{ij}]_{m\times n}$. Now, consider

$$\begin{aligned} v &= \mathscr{A}(u) \\ &= \mathscr{A}(\sum_{j=1}^{n} x_j u_j) = \sum_{j=1}^{n} x_j \mathscr{A}(u_j) \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} v_i \\ &= \sum_{i=1}^{m} v_i \left(\sum_{j=1}^{n} a_{ij} x_j \right), \end{aligned}$$

that is the coordinates of v with respect to basis $\mathcal{V}$ are given by the entries of the vector $\boldsymbol{Ax}$, where $\boldsymbol{x}$ is the vector of coordinates of u with respect to the basis $\mathcal{U}$, which proves the theorem. ■

Remark 10.24. Note that the theorem also provides a recipe to find $\boldsymbol{A}$: the columns of $\boldsymbol{A}$ are exactly the coordinates of the images of the vectors from $\mathcal{U}$ with respect to $\mathcal{V}$.

The isomorphisms $\mathbf{F}^n \longrightarrow U$ and $\mathbf{F}^n \longrightarrow V$ determined by two bases help to explain the relation between $\mathscr{A}$ and $\boldsymbol{A}$, which can be represented in the form the following *commutative diagram*:

$$\begin{array}{ccc} \mathbf{F}^n & \xrightarrow{\boldsymbol{A}} & \mathbf{F}^m \\ {\scriptstyle\mathcal{U}}\big\downarrow & & \big\downarrow{\scriptstyle\mathcal{V}} \\ U & \xrightarrow{\mathscr{A}} & V \end{array}$$

The meaning of the words "commutative diagram" is that going from $\mathbf{F}^n$ to V along any of the two possible paths in it will give the same answer.

Examples will be given in the next section.

10.5 Linear operators

Definition 10.25. *A linear transformation $\mathscr{A}$ from vector space V to itself is called a linear operator.*

Note that now, if we'd like to talk about the matrix of linear transformation, we would fix just one basis $\mathcal{B}$ of V. Moreover, both $\ker\mathscr{A}$ and $\operatorname{im}\mathscr{A}$ are subspaces of V.

Proposition 10.26. *Let $\boldsymbol{A}$ be the matrix of a linear operator $\mathscr{A}$ with respect to a basis $\mathcal{B}$. Suppose that the matrix of the basis change to a new basis $\mathcal{B}'$ is $\boldsymbol{P}$ (see the previous lecture for the definition). Then the matrix that represents $\mathscr{A}$ with respect to this new basis is $\boldsymbol{A}' = \boldsymbol{P}^{-1}\boldsymbol{AP}$.*

Proof. We just use what we already know. Let $v = \mathscr{A}(u)$. I have that $[v]_{\mathcal{B}'} = \boldsymbol{P}^{-1}[v]_{\mathcal{B}}$. Therefore

$$[v]_{\mathcal{B}'} = \boldsymbol{P}^{-1}[\mathscr{A}(u)]_{\mathcal{B}} = \boldsymbol{P}^{-1}[\mathscr{A}]_{\mathcal{B}}[u]_{\mathcal{B}} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}[u]_{\mathcal{B}'},$$

which proves the proposition. ■

Remark 10.27. Two matrices $\boldsymbol{A}$ and $\boldsymbol{A}'$ as in the last proposition are called *similar* matrices.

Example 10.28. Consider the space of polynomials $\mathbf{P}_3$ with degree at most three. We know that $\mathcal{B} = (1, x, x^2, x^3)$ is the standard basis of $\mathbf{P}_3$. Consider the linear transformation $\mathscr{D} = \frac{\mathrm{d}}{\mathrm{d}x}$ of differentiations of elements of $\mathbf{P}_3$:

$$\mathscr{D} \colon \mathbf{P}_3 \longrightarrow \mathbf{P}_3.$$

What is the matrix $\boldsymbol{D} = [\mathscr{D}]_{\mathcal{B}}$?

To figure out the matrix $\boldsymbol{D}$ we must see how $\mathscr{D}$ acts on the basis vectors:

$$\begin{aligned}
\mathscr{D}(1) &= 0 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + 0 \cdot x^3, \\
\mathscr{D}(x) &= 1 = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + 0 \cdot x^3, \\
\mathscr{D}(x^2) &= 2x = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 + 0 \cdot x^3, \\
\mathscr{D}(x^3) &= 3x^2 = 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2 + 0 \cdot x^3,
\end{aligned}$$

so that

$$\boldsymbol{D} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now define the functions $g_i(x) = (x+t)^{i-1}, i = 1, 2, 3, 4$ for any real $t \in \mathbf{R}$. Clearly these functions are in $\mathbf{P}_3$ and hence can be represented as linear combinations of vectors in $\mathcal{B}$:

$$\begin{aligned}
g_1(x) &= 1 = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + 0 \cdot x^3, \\
g_2(x) &= t + x = t \cdot 1 + 1 \cdot x + 0 \cdot x^2 + 0 \cdot x^3, \\
g_3(x) &= t^2 + 2tx + x^2 = t^2 \cdot 1 + 2t \cdot x + 1 \cdot x^2 + 0 \cdot x^3, \\
g_4(x) &= t^3 + 3t^2x + 3tx^2 + x^3 = t^3 \cdot 1 + 3t^2 \cdot x + 3t \cdot x^2 + 1 \cdot x^3.
\end{aligned}$$

Putting the coordinates of g_i into the matrix

$$\boldsymbol{P} = \begin{bmatrix} 1 & t & t^2 & t^3 \\ 0 & 1 & 2t & 3t^2 \\ 0 & 0 & 1 & 3t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

I can see that $\boldsymbol{P}$ is invertible with the inverse

$$\boldsymbol{P}^{-1} = \begin{bmatrix} 1 & -t & t^2 & -t^3 \\ 0 & 1 & -2t & 3t^2 \\ 0 & 0 & 1 & -3t \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

therefore I conclude that $\mathscr{B}' = (g_1, g_2, g_3, g_4)$ is another basis for $\mathbf{P}_3$.

Now I can calculate my matrix $\boldsymbol{D}'$ of $\mathscr{D}$ in the basis $\mathscr{B}'$. Some (tedious) computations give me

$$\boldsymbol{P}^{-1}\boldsymbol{D}\boldsymbol{P} = \boldsymbol{D},$$

and hence the matrix is the same (can you see a simpler way to conclude this?).

Anyway, now we are posed with a very important question: How to choose a basis of V to make sure that the matrix of a given linear transformation is simplest in this basis? Of course, we must be very clear to understand what we mean by "simplest." Basically the rest of the course will be devoted to get a (partial) answer to this question.