Linear Algebra Notes

Lecture Notes, University of Toronto, Fall 2016

1. Linear maps

(Ctd')

1.1. Isomorphisms.

Definition 1.1. An invertible linear map $T\colon V \to W$ is called a linear isomorphism from V to W.

Etymology: morphism=shape, iso=same.

Examples 1.2. (a) The map

$$T\colon \mathcal{P}_n(F) \to F^{n+1},\ p \mapsto (a_0, \ldots, a_n)$$

for $p(x) = \sum a_i x^i$ is an isomorphism. Similarly the map

$$T\colon \mathcal{P}_n(F) \to F^{n+1},\ p \mapsto (b_0, \ldots, b_n)$$

for $p(x) = \sum b_i (x+1)^i$.

(b) Given distinct elements $c_0, \ldots, c_n \in F$, the map

$$T\colon \mathcal{P}_n(F) \to F^{n+1},\ p \mapsto (p(c_0), \ldots, p(c_n))$$

is an isomorphism.

(c) Let V be the vector space of *finite* sequences. Then

$$\mathcal{P}(\mathbb{R}) \to V, \quad p \mapsto (a_0, \ldots, a_n, \ldots)$$

for $p(x) = \sum a_i x^i$ is an isomorphism. (Why? Lagrange interpolation!)

An isomorphism 'identifies' spaces (but the choice of identification depends on the choice of isomorphism). For instance, in the last example, the operator on polynomials given as 'multiplication by x', translates into a shift operator $(a_0, a_1, a_2 \ldots) \mapsto (0, a_0, a_1, \ldots)$.

Theorem 1.3. *If $T\colon V \to W$ is an isomorphism, then* $\dim V = \dim W$.

Proof. If both V, W are infinite-dimensional there is nothing to show. If $\dim V < \infty$, then

$$\dim(V) = \dim(N(T)) + \dim(R(T)) = 0 + \dim(W) = \dim(W)$$

by rank-nullity theorem. If $\dim W < \infty$, use the same argument for the isomorphism $T^{-1}\colon W \to V$. $\square$

Theorem 1.4. *Suppose $T\colon V \to W$ is linear, and that* $\dim(V) = \dim(W) < \infty$. *Then*

$$T \text{ is an isomorphism} \quad \Leftrightarrow \quad N(T) = \{0\} \quad \Leftrightarrow \quad R(T) = W.$$

Proof. We have $N(T) = \{0\} \Leftrightarrow \dim(N(T)) = 0$ and $R(T) = W \Leftrightarrow \dim(R(T)) = \dim(W)$. By rank-nullity, $\dim(W) = \dim(V) = \dim(N(T)) + \dim(R(T))$, it follows that T is onto if and only if T is one-to-one. $\square$

Note that finite-dimensionality is important here. For example,

$$T\colon \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R}),\ p \mapsto p'$$

is onto, but not one-to-one.

1.2. **The space of linear maps.** Let $\mathcal{L}(V, W)$ be the set of all linear maps from V to W. Define addition and scalar multiplication 'pointwise' by

$$(T_1 + T_2)(v) = T_1(v) + T_2(v), \quad (aT)(v) = a\,T(v).$$

for $T_1, T_2, T \in \mathcal{L}(V, W)$ and $a \in F$.

Theorem 1.5. *With these operations, $\mathcal{L}(V, W)$ is a vector space.*

The proof is a straightforward check, left as an exercise. (Actually, for any set X, the space $\mathcal{F}(X, U)$ of functions from X to a given vector space U is a vector space. $\mathcal{L}(V, W)$ is a vector subspace of $\mathcal{F}(V, W)$.) Important special cases:

- $\mathcal{L}(F, V)$ is identified with V, by the isomorphism $V \to \mathcal{L}(F, V)$ taking $v \in V$ to the linear map $F \to V,\ a \mapsto av$.
- The space $\mathcal{L}(V, F)$ is called the dual space to V, denoted V^*. Its elements are called the *linear functionals* on V. We'll return to this later.

Lemma 1.6. *If $T\colon V \to W$ and $S\colon U \to V$ are linear maps, then the composition $T \circ S\colon U \to V$ is linear.*

Proof.

$$T(S(u_1 + u_2)) = T(S(u_1) + S(u_2)) = T(S(u_1)) + T(S(u_2)),$$

and similarly

$$T(S(au)) = (T(aS(u))) = a(T(S(u))).$$

$\square$

Thus, we have a map

$$\mathcal{L}(V, W) \times \mathcal{L}(U, V) \to \mathcal{L}(U, W), \quad (T, S) \mapsto T \circ S.$$

Some simple properties of the composition:

$$(T_1 + T_2) \circ S = T_1 \circ S + T_2 \circ S, \quad (aT) \circ S = a\,T \circ S,$$

meaning it's linear in the first argument, and

$$T \circ (S_1 + S_2) = T \circ S_1 + T \circ S_2, \quad T \circ (aS) = a\,T \circ S,$$

meaning it's also also linear in the second argument. (The composition is an example of a *bilinear map.*)

For the special case $V = W$, we will also use the notation $\mathcal{L}(V) = \mathcal{L}(V, V)$. The map $I \in \mathcal{L}(V)$ given by $I(v) = v$ is called the *identity map* (also denoted I_V for emphasis). Note that elements $T \in \mathcal{L}(V)$ can be iterated:

$$T^k = T \circ \cdots \circ T,$$

for $k \geq 0$ with the convention $T^0 = I$. Note also that in general,

$$T \circ S \neq S \circ T.$$

Example 1.7. Let V be the vector space of infinitely-differentiable functions on $\mathbb{R}$. Then $\frac{\partial}{\partial x} \in \mathcal{L}(V)$. Finite linear combinations

$$T = \sum_{i=0}^{N} a_i (\frac{\partial}{\partial x})^i$$

are called *constant coefficient differential operators.*

Theorem 1.8. *A linear map $T \in \mathcal{L}(V, W)$ is invertible if and only if there exists a linear map $S \in \mathcal{L}(W, V)$ with $S \circ T = I_V$ and $T \circ S = I_W$.*

Proof. One direction is obvious: if T is invertible, then $S = T^{-1}$ has the properties stated. For the converse, if S with $S \circ T = I_V$ and $T \circ S = I_W$ is given, then $N(T) = \{0\}$ (since $T(v) = 0$ implies $v = S(T(v)) = S(0) = 0$), and $R(T) = W$ (since $w = T(S(w))$ for all $w \in W$). $\square$

Note that, a bit more generally, it suffices to have two maps $S_1, S_2 \in \mathcal{L}(W, V)$ with $S_1 \circ T = I_V$ and $T \circ S_2 = I_W$. It is automatic that $S_1 = S_2$, since

$$S_1 = S_1 \circ I_W = S_1 \circ (T \circ S_2) = (S_1 \circ T) \circ S_2 = I_V \circ S_2 = S_2.$$

in that case.

Remark 1.9. (a) If $\dim V = \dim W < \infty$, then the conditions $S \circ T = I_V$ and $T \circ S = I_W$ are equivalent. Indeed, the first condition implies $N(T) = \{0\}$, hence $R(T) = W$ by an earlier theorem.

(b) If $\dim V \neq \dim W$ then **both** conditions are needed. For example,

$$T \colon \mathbb{R} \to \mathbb{R}^2,\ t \mapsto (t, 0), \quad S \colon \mathbb{R}^2 \to \mathbb{R},\ (t_1, t_2) \mapsto t_1$$

satisfiies $S \circ T = I$, but neither S nor T are invertible.

(c) If $\dim V = \dim W = \infty$, then **both** conditions are needed. For example, let $S, T \colon F^\infty \to F^\infty$ where

$$T(a_0, a_1, a_2 \ldots) = (0, a_0, a_1, \ldots), \qquad S(a_0, a_1, a_2 \ldots) = (a_1, a_2, a_3, \ldots)$$

Then $S \circ T$ is the identity but $T \circ S$ is not.

2. Matrix representations of linear maps

2.1. **Matrix representations.** Let V be a vector space with $\dim V = n < \infty$, and let $\beta = \{v_1, \ldots, v_n\}$ be an *ordered* basis. By this, we mean that the basis vectors have been enumerated; thus it is more properly an ordered list $v_1, \ldots, v_n$. For $v = a_1 v_1 + \ldots + a_n v_n \in V$, write

$$[v]_\beta = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

This is called the *coordinate vector* of v in the basis β. The map

$$\varphi_\beta \colon V \to F^n, \quad v \mapsto [v]_\beta$$

is a linear isomorphism, with inverse the map

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \mapsto a_1 v_1 + \ldots + a_n v_n.$$

(Well, we really defined φ_β in terms of the inverse.)

Example 2.1. Let $V = \mathcal{P}_3(\mathbb{R})$ with its standard basis $1, x, x^2, x^3$. For $p = 2x^3 - 3x^2 + 1$ we have

$$[v]_\beta = \begin{pmatrix} 1 \\ 0 \\ -3 \\ 2 \end{pmatrix}.$$

The choice of a basis '*identifies*' V with F^n. Of course, the identification depends on the choice of basis, and later we will investigate this dependence.

What about linear maps? Let $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_m\}$ be bases of V and W, respectively. Each $T(v_j)$ is a linear combination of the w_i's:

$$T(v_j) = A_{1j} w_1 + \ldots + A_{mj} w_m.$$

(Beware the ordering; we're writing A_{1j} not A_{j1} etc.) The coefficients define a matrix,

$$[T]_\beta^\gamma = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix}$$

If $W = V$ and $\gamma = \beta$, we also write $[T]_\beta = [T]_\beta^\beta$.

Note:

$$\boxed{\textit{The coefficients of } T(v_j) \textit{ form the } j\textit{-th } \text{column} \textit{ of } [T]_\beta^\gamma.}$$

In other words, the j-th column of $[T]_\beta^\gamma$ is $[T(v_j)]_\gamma$.

Example 2.2. Let $V = \mathcal{P}_3(\mathbb{R})$, $W = \mathcal{P}_2(\mathbb{R})$ and $T(p) = p' + p'''$. Let us find $[T]_\beta^\gamma$ in terms of the standard bases $\beta = \{1, x, x^2, x^3\}$ and $\gamma = \{1, x, x^2\}$. Since

$$T(1) = 0,\ T(x) = 1,\ T(x^2) = 2x,\ T(x^3) = 3x^2 + 6$$

we find

$$[T]_\beta^\gamma = \begin{pmatrix} 0 & 1 & 0 & 3 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 6 \end{pmatrix}$$

Example 2.3. The identity map $I\colon V \to V$ has the coordinate expression, for any ordered basis β,

$$[I]_\beta^\beta = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

This is the 'identity matrix'. (Note however that for *distinct* bases β, γ, the matrix $[I]_\beta^\gamma$ will look different.)

Example 2.4. If $T\colon V \to W$ is an isomorphism, and $\beta = \{v_1, \ldots, v_n\}$ is a basis of V, then $w_i = T(v_i)$ form a basis $\gamma = \{w_1, \ldots, w_n\}$ of W. In terms of these bases, $[T]_\beta^\gamma$ is the identity matrix.

Theorem 2.5. *Let V, W be finite-dimensional vector spaces with bases β, γ. The map*

$$\mathcal{L}(V, W) \to M_{m\times n}(F), \quad T \mapsto [T]_\beta^\gamma$$

is an isomorphism of vector spaces. In particular,

$$\dim(\mathcal{L}(V, W)) = \dim(V)\dim(W).$$

Proof. Since $[T_1 + T_2]_\beta^\gamma = [T_1]_\beta^\gamma + [T_2]_\beta^\gamma$ and $[aT]_\beta^\gamma = a[T]_\beta^\gamma$, the map is linear.

It is injective (one-to-one), since $[T]_\beta^\gamma = 0$ means that $[T(v_i)]_\gamma = 0$ for all i, hence $T(v_i) = 0$ for all i, hence $T(v) = 0$ for all $v = \sum a_i v_i \in V$.

It is surjective (onto) since every $A \in M_{m\times n}(F)$, with matrix entries A_{ji}, defines a linear transformation by

$$T(\sum a_i v_i) = \sum_{ij} A_{ji} a_i w_j,$$

with $[T]_\beta^\gamma = A$.

The dimension formula follows from $\dim(M_{m\times n}(F)) = mn$. $\square$

Suppose $T(v) = w$. What is the relationship between $[v]_\beta$, $[w]_\gamma$, and $[T]_\beta^\gamma$? If $v = \sum a_j v_j \in V$ and $w = \sum b_i w_i \in W$, then

$$\begin{aligned} T(v) &= T(\sum_{i=1}^n a_j v_j) \\ &= \sum_{j=1}^n a_j T(v_j) \\ &= \sum_{j=1}^n a_j \sum_{i=1}^m A_{ij} w_i \\ &= \sum_{i=1}^m \Big(\sum_{j=1}^n A_{ij} a_j\Big) w_i \end{aligned}$$

Hence

$$b_i = \sum_{j=1}^n A_{ij} a_j,$$

or on full display:

$$\begin{aligned} b_1 &= A_{11}a_1 + \ldots + A_{1n}a_n \\ b_2 &= A_{21}a_1 + \ldots + A_{2n}a_n \\ \cdots &= \ \cdots \\ b_m &= A_{m1}a_1 + \ldots + A_{mn}a_n \end{aligned}$$

We write this in *matrix notation*

$$\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

In short,

$$\boxed{[w]_\gamma = [T]_\beta^\gamma \ [v]_\beta.} \tag{1}$$

2.2. **Composition.** Let $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(U, V)$ be linear maps, and $T \circ S \in \mathcal{L}(U, W)$ the composition. Let $\alpha = \{u_1, \ldots, u_l\}, \beta = \{v_1, \ldots, v_n\}, \gamma = \{w_1, \ldots, w_m\}$ be ordered bases of U, V, W, respectively. We are interested in the relationship between the matrices

$$A = [T]_\beta^\gamma, \quad B = [S]_\alpha^\beta, \quad C = [T \circ S]_\alpha^\gamma.$$

We have

$$T(S(u_k)) = \sum_{j=1}^{m} T(B_{jk}v_j) = \sum_{i=1}^{m}\sum_{j=1}^{n} B_{jk}A_{ij}w_i.$$

Hence,

$$C_{ik} = \sum_{j=1}^{m} A_{ij}B_{jk}.$$

We take the expression on the right hand side as the definition of $(AB)_{ik}$:

$$(AB)_{ik} := \sum_{j=1}^{m} A_{ij}B_{jk}$$

It defines a matrix multiplication,

$$M_{m\times n}(F) \times M_{n\times l}(F) \to M_{l\times n}(F), \quad (A, B) \mapsto AB$$

where

$$\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{11} & \cdots & B_{1l} \\ \vdots & & \vdots \\ B_{n1} & \cdots & B_{nl} \end{pmatrix} = \begin{pmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & & \vdots \\ C_{l1} & \cdots & C_{ln} \end{pmatrix}$$

with $C_{ik} = \sum_{j=1}^{m} A_{ij}B_{jk}$. **Note:**

$$\boxed{\text{The matrix entry } C_{ik} \text{ is the product of the } i\text{-th row of } A \text{ with the } j\text{-th column of } B.}$$

Example 2.6. An example with numbers:

$$\begin{pmatrix} 1 & -3 & 4 & 3 \\ 0 & 4 & 7 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 & -1 \\ 2 & -2 & 0 \\ 0 & 0 & -1 \\ 1 & 4 & 3 \end{pmatrix} = \begin{pmatrix} -2 & 20 & 4 \\ 7 & -12 & -10 \end{pmatrix}$$

Example 2.7. Another example:

$$\begin{pmatrix} -2 & 20 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \end{pmatrix} = (10), \quad \begin{pmatrix} 5 \\ 1 \end{pmatrix} \begin{pmatrix} -2 & 20 \end{pmatrix} = \begin{pmatrix} -10 & 100 \\ -2 & 20 \end{pmatrix}$$

In any case, with this definition of matrix multiplication we have that

$$\boxed{[T \circ S]^\gamma_\alpha = [T]^\gamma_\beta \, [S]^\beta_\alpha.} \tag{2}$$

In summary, the choice of bases identifies linear transformations with matrices, and composition of linear maps becomes matrix multiplication.

Remark 2.8. We could also say it as follows: Under the identification

$$\mathcal{L}(F^n, F^m) \cong M_{m\times n}(F),$$

the composition of operators $F^l \xrightarrow{S} F^n \xrightarrow{T} F^m$ corresponds to matrix multiplication.

Mathematicians like to illustrate this with *commutative diagrams*, as follows: [...]

Some remarks:

1. The action of a matrix on a vector is a special case of matrix multiplication, by viewing a column vector as an $n \times 1$-matrix.
2. Using the standard bases for F^n, F^m we have the isomorphism
$$\mathcal{L}(F^n, F^m) \to M_{m\times n}(F).$$
Suppose the matrix A corresponds to the linear map T. Writing the elements of F^n, F^m as column vectors, the j-column of A is the image of the j-th standard basis vector under T. [Example...] In the opposite direction, every $A \in M_{m\times n}(F)$ determines a linear map. We denote this by
$$L_A \colon F^n \to F^m.$$
Writing the elements of F^n as column vectors, this is just the matrix multiplication from the left (hence the notation).
3. Since matrix multiplication is just a special case of composition of linear maps, it's clear that it is linear in both arguments:
$$(A_1 + A_2)B = A_1B + A_2B, \quad (aA)B = a(AB),$$
$$A(B_1 + B_2) = AB_1 + AB_2, \quad A(aB) = a(AB).$$
4. Taking $n = m$, we have addition and multiplication on the space $M_{n\times n}(F)$ of square matrices. There is also an 'additive unit' $\mathbf{0}$ given by the zero matrix, and a 'multiplicative unit' $\mathbf{1}$ given by the identity matrix (with matrix entries $\mathbf{1}_{ij} = \delta_{ij}$, equal to 1 if $i = j$ and zero otherwise). If $n = 1$, these are just the usual addition and multiplication

of $M_{1\times 1}(F)=F$. But if $n\geq 2$, it violates several of the field axioms. First of all, the product is non-commutative:

$$AB\neq BA$$

in general. Secondly, $A\neq \mathbf{0}$ does **not** guarantee the existence of an inverse. As a consequence, $AB=\mathbf{0}$ does *not* guarantee that $A=0$ or $B=0$. (One even has examples with $A^2=\mathbf{0}$ but $A\neq\mathbf{0}$ – can you think of one?)

2.3. **Change of bases.** As was mentioned before, we have to understand the dependence of all constructions on the choice of bases. Let β,β' be two ordered bases of V, and $[v]_\beta,\ [v]_{\beta'}$ the expression in the two bases. What's the relation between these coordinate vectors?

Actually, we 'know' the answer already, as a special case of Equation (1) – the formula $(T(v))_{[\gamma]}=[T]^\gamma_\beta[v]_\beta$ tells us that:

$$[v]_{\beta'}=[I_V]^{\beta'}_\beta[v]_\beta,\quad [v]_\beta=[I_V]^\beta_{\beta'}[v]_{\beta'}.$$

(matrix multiplication) where I_V is the identity. We call

$$[I_V]^\beta_{\beta'}$$

the *change of coordinate matrix*, for changing the basis from β' to β. The change-of coordinate matrix in the other direction is the inverse:

$$\big([I_V]^\beta_{\beta'}\big)^{-1}=[I_V]^{\beta'}_\beta.$$

This follows from the action on coordinate vectors, or alternatively from

$$[I_V]^\beta_{\beta'}\circ[I_V]^{\beta'}_\beta=[I_V\circ I_V]^\beta_\beta=[I_V]^\beta_\beta=\mathbf{1}.$$

Example 2.9. Let $V=\mathbb{R}^2$. Given an angle θ, consider the basis

$$\beta=\{(\cos\theta,\sin\theta),(-\sin\theta,\cos\theta)\}.$$

of $V=\mathbb{R}^2$. Given another angle θ', we get a similar basis β'. What is the change of basis matrix? Since β,β' are obtained from the standard basis of $\mathbb{R}^2$ by rotation by θ,θ', one expects that the change of basis matrix should be a rotation by $\theta'-\theta$.

Indeed, using the standard trig identities,

$$\begin{aligned}\cos\alpha\ (\cos\theta,\sin\theta)+\sin\alpha\ (-\sin\theta,\cos\theta)\\ =\big(\cos(\theta+\alpha),\sin(\theta+\alpha)\big),\\ -\sin\alpha\ (\cos\theta,\sin\theta)+\cos\alpha\ (-\sin\theta,\cos\theta)\\ =\big(-\sin(\theta+\alpha),\cos(\theta+\alpha)\big).\end{aligned}$$

Thus, taking $\alpha=\theta'-\theta$, this expresses the basis β' in terms of β. We read off the columns of $(I_V)^\beta_{\beta'}$ as the coefficients for this change of bases:

$$(I_V)^\beta_{\beta'}=\begin{pmatrix}\cos(\theta'-\theta) & -\sin(\theta'-\theta)\\ \sin(\theta'-\theta) & \cos(\theta'-\theta)\end{pmatrix}.$$

Example 2.10. Suppose $c_0, \ldots, c_n$ are distinct. Let $V = \mathcal{P}_n(F)$, let

$$\beta = \{1, x, \ldots, x^n\}$$

be the standard basis, and

$$\beta' = \{p_0, \ldots, p_n\}$$

the basis given by the Lagrange interpolation polynomials. What is the change-of-coordinate matrix $[I_V]_\beta^{\beta'}$? Recall that if p is given, then

$$p(x) = \sum_i p(c_i)\ p_i(x)$$

by the Lagrange interpolation. We can use this to express the standard basis in terms of the Lagrange basis.

$$x^k = \sum_i c_i^k\ p_i(x).$$

The coefficients of the β-basis vectors in terms of the β'-basis vectors form the columns of the change-of-basis matrix from β to β'. Hence, $Q := [I_V]_\beta^{\beta'}$ is the matrix,

$$Q = \begin{pmatrix} 1 & c_0 & \cdots & c_0^n \\ 1 & c_1 & \cdots & c_1^n \\ \vdots & \vdots & & \vdots \\ 1 & c_n & \cdots & c_n^n \end{pmatrix}$$

The matrix for the other direction $[I_V]_{\beta'}^{\beta}$ is less straightforward, as it amounts to computing the coefficients of the Lagrange interpolation polynomials in terms of the standard basis. By the way, note what we've stumbled upon the following fact:

> *The coefficients of the Lagrange interpolation polynomials are obtained as the columns of the matrix Q^{-1}, where Q is the matrix above.*

Let us next consider how the matrix representations of operators $T \in \mathcal{L}(V)$ change with the change of basis. We write $[T]_\beta = [T]_\beta^\beta$.

Proposition 2.11. *Let V be a finite-dimensional vector space, and β, β' two ordered bases. Then the change of coordinate matrix Q (from β to β') is invertible, with inverse Q^{-1} the change of coordinate matrix from β' to β. We have*

$$[v]_{\beta'} = Q\ [v]_\beta.$$

for all $v \in V$. Given a linear map $T \in \mathcal{L}(V)$, we have

$$[T]_{\beta'} = Q\ [T]_\beta\ Q^{-1}.$$

Proof. The formula $[v]_{\beta'} = Q\ [v]_\beta$ holds by the calculation above. Since

$$[I_V]_{\beta'}^{\beta}[I_V]_\beta^{\beta'} = [I_V]_\beta^\beta, \qquad [I_V]_\beta^{\beta'}[I_V]_{\beta'}^{\beta} = [I_V]_{\beta'}^{\beta'}$$

are both the identity matrix, we see that the change of coordinate matrix from β' to β is the inverse matrix:

$$[I_V]_{\beta'}^{\beta} = Q^{-1}.$$

Similarly, using (2),

$$[T]^{\beta'}_{\beta'} = [I_V]^{\beta'}_{\beta}\ [T]^{\beta}_{\beta}\ [I_V]^{\beta}_{\beta'} = Q\ [T]_{\beta}\ Q^{-1}.$$

□

Definition 2.12. To matrices $A, A' \in \mathrm{Mat}_{n\times n}$ are called *similar* if there exists an invertible matrix Q such that

$$A' = Q\, A\, Q^{-1}.$$

Thus, we see that the matrix representations of $T \in \mathcal{L}(V)$ in two different bases are related by a similarity transformation.

3. Dual spaces

Given a vector space V, one can consider the space of linear maps $\phi\colon V \to F$. Typical examples include:

- For the vector space of functions from a set X to F, and any given $c \in X$, the evaluation
$$\mathrm{ev}_c\colon \mathcal{F}(F,F) \to F, \quad f \mapsto f(c).$$
- For the vector space F^n, written as column vectors, the i-th coordinate function
$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto x_i.$$
More generally, any given $b_1, \ldots, b_n$ defines a linear functional
$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto x_1 b_1 + \ldots + x_n b_n.$$
(Note that this can also be written as matrix multiplication with the row vector $(b_1, \ldots, b_n)$.)
- The *trace* of a matrix,
$$\mathrm{tr}\colon V = \mathrm{Mat}_{n\times n}(F) \to F,\ A \mapsto A_{11} + A_{22} + \ldots + A_{nn}.$$
More generally, for a fixed matrix $B \in \mathrm{Mat}_{n\times n}(F)$, there is a linear functional
$$A \mapsto \mathrm{tr}(BA).$$

Definition 3.1. For any vector space V over a field F, we denote by

$$V^* = \mathcal{L}(V, F)$$

the *dual space* of V.

Note that

$$\dim(V^*) = \dim \mathcal{L}(V,F) = \dim(V)\dim(F) = \dim(V).$$

(This also holds true if V is infinite dimensional.)

Remark 3.2. If V is finite-dimensional, this means that V and V^* are isomorphic. But this is false if $\dim V = \infty$. For instance, if V has an infinite, but countable basis (such as the space $V = \mathcal{P}(F)$), one can show that V^* does *not* have a countable basis, and hence cannot be isomorphic to V.

Suppose $\dim V < \infty$, and let $\beta = \{v_1, \dots, v_n\}$ be a basis of V. Then any linear functional $f \colon V \to F$ is determined by its values on the basis vectors: Given $b_1, \dots, b_n \in F$, we obtain a unique linear functional such that

$$f(v_i) = b_i, \quad i = 1, \dots, n.$$

Namely, $f(a_1 v_1 + \ldots + a_n v_n) = a_1 b_1 + \ldots + a_n b_n$. In particular, given j there is a unique linear functional taking on the value 1 on v_j and the value 0 on all other basis vectors. This linear functional is denoted v_j^*:

$$v_j^*(v_i) = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Proposition 3.3. *The set $\beta^* = \{v_1^*, \dots, v_n^*\}$ is a basis of V^*, called the* dual basis.

Proof. Since $\dim(V^*) = \dim(V) = n$, it suffices to show that the v_i^* are linearly independent. Thus suppose

$$\lambda_1 v_1^* + \cdots + \lambda_n v_n^* = 0.$$

Evaluating on v_i, and using the definition of dual basis, we get $\lambda_i = 0$. □

Example 3.4.

Definition 3.5. Suppose V, W are vector spaces, and V^*, W^* their dual spaces. Given a linear map

$$T \colon V \to W$$

define a *dual map* (or *transpose map*) $T^t \colon W^* \to V^*$ as follows: If $\psi \colon W \to F$ is a linear functional, then

$$T^t(\psi) = \psi \circ T \colon V \to F.$$

Note that T^t is a linear map from W^* to V^* since

$$T^t(\psi_1 + \psi_2) = (\psi_1 + \psi_2) \circ T = \psi_1 \circ T + \psi_2 \circ T;$$

similarly for scalar multiplication. Note also that the dual map goes in the 'opposite direction'. In fact, under composition,

$$(T \circ S)^t = S^t \circ T^t.$$

(We leave this as an Exercise.)

Theorem 3.6. *Let β, γ be ordered bases for V, W respectively, and β^*, γ^* the dual bases of V^*, W^* respectively. Then the matrix of T^t with respect to the dual bases β^*, γ^* of V^*, W^* is the transpose of the matrix $[T]_\beta^\gamma$:*

$$[T^t]_{\gamma^*}^{\beta^*} = \left([T]_\beta^\gamma\right)^t.$$

Proof. Write $\beta = \{v_1, \dots, v_n\},\ \gamma = \{w_1, \dots, w_m\}$ and $\beta^* = \{v_1^*, \dots, v_n^*\},\ \gamma = \{w_1^*, \dots, w_m^*\}$.
Let $A = [T]_\beta^\gamma$ be the matrix of T. By definition,

$$T(v_j) = \sum_k A_{kj} w_k.$$

Similarly, the matrix $B = [T^t]_{\gamma^*}^{\beta^*}$ is defined by

$$(T^t)(w_i^*) = \sum_l B_{li} v_l^*.$$

Applying w_i^* to the first equation, we get

$$w_i^*(T(v_j)) = \sum_k A_{kj} w_i^*(w_k) = A_{ij}.$$

On the other hand,

$$w_i^*(T(v_j)) = (T^t(w_i^*))(v_j) = \sum_l B_{li} v_l^*(v_j) = B_{ji}.$$

This shows $A_{ij} = B_{ji}$. □

This shows that the dual map is the 'coordinate-free' version of the transpose of a matrix.

4. Elementary matrix operations

4.1. **The Gauss elmination in terms of matrices.** Consider a system of linear equations, such as,

$$\begin{aligned} -2y + z &= 5 \\ x - 4y + 6z &= 10 \\ 4x - 11y + 11z &= 12 \end{aligned}$$

The standard method of solving such a system is *Gauss elimination.* We first interchange the first and second equation (since we'd like to have an x in the first equation).

$$\begin{aligned} x - 4y + 6z &= 10 \\ -2y + z &= 5 \\ 4x - 11y + 11z &= 12 \end{aligned}$$

Next, we subtract 4 times the first equation from the last (to get rid of x's in all but the first equation),

$$\begin{aligned} x - 4y + 6z &= 10 \\ -2y + z &= 5 \\ 5y - 13z &= -28 \end{aligned}$$

Now, divide the second equation by -2,

$$\begin{aligned} x - 4y + 6z &= 10 \\ y - \frac{1}{2}z &= -\frac{5}{2} \\ 5y - 13z &= -28 \end{aligned}$$

and subtract suitable multiples of it from the first and third equation, to arrive at

$$\begin{aligned} x + 4z &= 0 \\ y - \frac{1}{2}z &= -\frac{5}{2} \\ -\frac{21}{2}z &= -\frac{31}{2} \end{aligned}$$

Finally, multiply the last equation by $-\frac{2}{21}$,

$$\begin{aligned} x + 4z &= 0 \\ y - \frac{1}{2}z &= -\frac{5}{2} \\ z &= -\frac{31}{21} \end{aligned}$$

and subtract suitable multiples of it from the second and third equation:

$$\begin{aligned} x &= -\frac{124}{21} \\ y &= -\frac{37}{21} \\ z &= -\frac{31}{21} \end{aligned}$$

What we used here is the fact that for a given system of equations

$$\begin{aligned} A_{11}x_1 + \ldots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + \ldots + A_{2n}x_n &= b_2 \\ &\vdots \\ A_{m1}x_1 + \ldots + A_{mn}x_n &= b_n \end{aligned}$$

the following three operations don't change the solution sets:

- interchange of two equations
- multiplying an equation by a non-zero scalar
- adding a scalar multiple of one equation to another equation

It is convenient to write the equation in matrix form

$$\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

The three operations above correspond to the *elementary row operations* on A, namely

- interchange of two rows
- multiplying a row by a non-zero scalar
- adding a scalar multiple of one row to another row

Of course, we have to apply the same operations to the right hand side, the b-vector. For this, it is convenient to work with the *augmented matrix*

$$\left(\begin{array}{ccc|c} A_{11} & \cdots & A_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ A_{m1} & \cdots & A_{mn} & b_m \end{array}\right)$$

and perform the elementary row operations on this larger matrix. One advantage of this method (aside from not writing the x_i's) is that we can at the same time solve several such equations with different right hand sides; one just has to augment one column for each right hand side:

$$\left(\begin{array}{ccc|cc} A_{11} & \cdots & A_{1n} & b_1 & c_1 \\ \vdots & & \vdots & \vdots & \vdots \\ A_{m1} & \cdots & A_{mn} & b_m & c_m \end{array}\right)$$

In the concrete example given above, the augmented matrix reads as

$$\left(\begin{array}{ccc|c} 0 & -2 & 1 & 5 \\ 1 & -4 & 6 & 10 \\ 4 & -11 & 11 & 12 \end{array}\right)$$

the first step was to exchange two rows:

$$\left(\begin{array}{ccc|c} 1 & -4 & 6 & 10 \\ 0 & -2 & 1 & 5 \\ 4 & -11 & 11 & 12 \end{array}\right)$$

Subtract 4 times the first row from the second row,

$$\left(\begin{array}{ccc|c} 1 & -4 & 6 & 10 \\ 0 & -2 & 1 & 5 \\ 0 & 5 & -13 & -28 \end{array}\right)$$

and so on. Eventually we get an augmented matrix of the form

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & -\frac{124}{21} \\ 0 & 1 & 0 & -\frac{37}{21} \\ 0 & 0 & 1 & -\frac{31}{21} \end{array}\right)$$

which tells us the solution, $x = -\frac{124}{21}$ and so on.

We will soon return to the theoretical and computational aspects of Gauss elimination. At this point, we simply took it as a motivation for considering the row operations.

Proposition 4.1. *If $A' \in M_{m\times n}(F)$ is obtained from $A \in M_{m\times n}(F)$ by an elementary row operation, then*

$$A' = PA$$

for a suitable invertible matrix $P \in M_{m\times m}(F)$. In fact, P is obtained by applying the given row operation to the identity matrix $I_m \in M_{m\times m}(F)$.

Idea of proof: The row operations can be interpreted as changes of bases for F^n. That is, letting $T = L_A \colon V = F^n \to W = F^m$ be the linear map defined by A,

$$A = [T]_\beta^\gamma$$

where β, γ are the standard bases, then

$$A' = [T]_\beta^{\gamma'}$$

for some new ordered basis γ'. So the formula is a special case of

$$[T]_\beta^{\gamma'} = [I_W]_\gamma^{\gamma'}[T]_\beta^\gamma.$$

where $P = [I_W]_\gamma^{\gamma'}$ is the change of basis matrix.

We won't spell out more details of the proof, but instead just illustrate it for with some examples where $m = 2, n = 3$. Consider the matrix

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix}$$

Interchange of rows:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix} = \begin{pmatrix} A_{21} & A_{22} & A_{23} \\ A_{11} & A_{12} & A_{13} \end{pmatrix}.$$

Multiplying the first row by non-zero scalar a:

$$\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix} = \begin{pmatrix} a\ A_{11} & a\ A_{12} & a\ A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix}.$$

Adding c-th multiple of first row to second row:

$$\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} + cA_{11} & A_{22} + cA_{12} & A_{23} + cA_{13} \end{pmatrix}.$$

Similar to the elementary row operations, we can also consider elementary column operations on matrices $A \in M_{m \times n}(F)$,

- interchange of two columns
- multiplying a column by a scalar
- adding a scalar multiple of one column to another column

Proposition 4.2. *Suppose A' is obtained from $A \in M_{m\times n}(F)$ by an elementary column operation. Then $A' = AQ$ where $Q \in M_{n\times n}$ is an invertible matrix. In fact, Q is obtained by applying the given column operation to the identity matrix.*

Again, Q is the change of basis matrix for three elementary types of base changes of F^n.

Warning: Performing column operations on the matrix for a system of equations does change the solution set! It should really be thought of as change of coordinates of the x_i's.

4.2. **The rank of a matrix.** Recall that the rank of a linear transformation $T\colon V \to W$ is the dimension of the range,
$$\operatorname{rank}(T) = \dim(R(T)).$$
Some observations: If $S\colon V' \to V$ is an isomorphism, then $\operatorname{rank}(T \circ S) = \operatorname{rank}(T)$, since $R(T \circ S) = R(T)$. If $U\colon W \to W'$ is an isomorphism, then $\operatorname{rank}(U \circ T) = \operatorname{rank}(T)$, since U restricts to an isomorphism from $R(T)$ to $R(U \circ T)$. Putting the two together, we have
$$\operatorname{rank}(U \circ T \circ S) = \operatorname{rank}(T)$$
if both U, S are linear isomorphisms.

Definition 4.3. For $A \in \mathrm{M}_{m\times n}(F)$, we define the *rank* of A to be the rank of the linear map $L_A\colon F^n \to F^m$ defined by A.

An easy consequence of this definition is that for any linear map $T\colon V \to W$ between finite-dimensional vector spaces, the rank of T coincides with the rank of $[T]_\beta^\gamma$ for any ordered bases β, γ.

Theorem 4.4. *The rank of $A \in M_{m\times n}(F)$ is the maximum number of linearly independent columns of A.*

Proof. Note that since the column vectors of A are the images of the standard basis vectors, the range of L_A is the space spanned by the columns. Hence, the rank is the dimension of the space spanned by the columns. But the dimension of the space spanned by a subset of a vector space is the cardinality of the largest linearly independent subset. □

Example 4.5. The matrices
$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \end{pmatrix}$$
have rank $1, 2, 2$ respectively.

Since we regard matrices as special cases of linear maps, we have that
$$\operatorname{rank}(PAQ) = \operatorname{rank}(A)$$
whenever P, Q are both invertible. In particular, we have:

Proposition 4.6. *The rank of a matrix is unchanged under elementary row or column operations.*

Proof. As we saw above, these operations are multiplications by invertible matrices, from the left or from the right. □

In examples, we can often use it to compute ranks of matrices rather quickly. Roughly, the strategy is to use the row-and columns operations to produce as many zeroes as possible.

Proposition 4.7. *Using elementary row and column operations, any matrix $A \in M_{m\times n}(F)$ of rank r can be put into 'block form'*
$$\begin{pmatrix} \mathbf{I_{r\times r}} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}$$

where the upper left block is the $r \times r$ identity matrix, and the other blocks are zero matrices of the indicated size.

Proof. Start with a matrix A. If $r > 0$, then A is non-zero, so it has at least one non-zero entry. Using interchanges of columns, and interchanges of rows, we can achieve that the entry in position $1,1$ is non-zero. Having achieved this, we divide the 1st row by this 1,1 entry, to make the 1,1 entry equal to 1. Subtracting multiples of the first row from the remaining rows, and then multiples of the first column from the remaining columns, we can arrange that the only entry in the first row or column is the entry $A_{11} = 1$. If there are further non-zero entries left, use interchanges of rows with row index > 1, or columns with column index > 1, to make the A_{22} entry non-zero. Divide by this entry to make it 1, and so forth. $\square$

Example 4.8.

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Note the following consequence of this result:

Proposition 4.9. *For every matrix A, we have that*
$$\operatorname{rank}(A) = \operatorname{rank}(A^t).$$
More generally, for any linear map $T\colon V \to W$ between finite-dimensional vector spaces we have $\operatorname{rank}(T) = \operatorname{rank}(T^t)$.