October 3, 2014

III. THE DUAL SPACE

BASED ON NOTES BY RODICA D. COSTIN

Contents

1. The dual of a vector space 1
1.1. Linear functionals 1
1.2. The dual space 2
1.3. Dual basis 3
1.4. Linear functionals as covectors; change of basis 4
1.5. Functionals and hyperplanes 5
1.6. Application: Spline interpolation of sampled data (*Based on Prof. Gerlach's notes)* *5*
1.7. The bidual 8
1.8. Orthogonality 8
1.9. The transpose 9
1.10. Eigenvalues of the transpose 11

1. The dual of a vector space

1.1. **Linear functionals.** Let V be a vector space over the scalar field $F = \mathbb{R}$ or $\mathbb{C}$.

Recall that linear functionals are particular cases of linear transformation, namely those whose values are in the scalar field (which is a one dimensional vector space):

Definition 1. *A* **linear functional** *on V is a linear transformation with scalar values: $\phi : V \to F$.*

A simple example is the first component functional: $\phi(\mathbf{x}) = x_1$.

We denote $\phi(\mathbf{x}) \equiv (\boldsymbol{\phi}, \mathbf{x})$.

Example in finite dimensions. If M is a row matrix, $M = [a_1, \ldots, a_n]$ where $a_j \in F$, then $\phi : F^n \to F$ defined by matrix multiplication, $\phi(\mathbf{x}) = M\mathbf{x}$, is a linear functional on F^n. These are, essentially, all the linear functionals on a finite dimensional vector space.

Indeed, the matrix associated to a linear functional $\phi : V \to F$ in a basis $\mathcal{B}_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V, and the (standard) basis $\{1\}$ of F is the row vector

$$[\phi_1, \ldots, \phi_n], \quad \text{where } \phi_j = (\boldsymbol{\phi}, \mathbf{v}_j) \tag{1}$$

If $\mathbf{x} = \sum_j x_j \mathbf{v}_j \in V$ then $(\boldsymbol{\phi}, \mathbf{x}) = (\boldsymbol{\phi}, \sum_j x_j \mathbf{v}_j) = \sum_j x_j (\boldsymbol{\phi}, \mathbf{v}_j) = \sum_j x_j \phi_j$ hence

$$(\boldsymbol{\phi}, \mathbf{x}) = [\phi_1, \ldots, \phi_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{2}$$

Examples in infinite dimensions. In infinite dimensions, linear functionals are also called *linear forms*. We mention a common notation in th Hilbert spaces used in quantum mechanics: *Dirac's bracket notation*

$$(\boldsymbol{\phi}, \mathbf{x}) \equiv \langle \boldsymbol{\phi} | \mathbf{x} \rangle$$

In this notation, a functional ϕ is rather denoted $\langle \boldsymbol{\phi} |$ and called *bra-*, while a vector $\mathbf{x} \in V$ is rather denoted $| \mathbf{x} \rangle$ and called *-ket*; they combine to $\langle \boldsymbol{\phi} | \mathbf{x} \rangle$, a *bra-ket*.

1. The most often encountered linear functionals are integrals. For example, on the linear space of continuous functions on $[a, b]$ let $I : C[a, b] \to \mathbb{R}$ defined as $(I, f) = \int_a^b f(t)\, dt$.

More generally, given some $w(t) > 0$ on $[a, b]$, the integral with respect to the weight w, $(I_w, f) = \int_a^b f(t)\, w(t) dt$ is a linear functional on $C[a, b]$.

2. Evaluation at a point: if $t_1 \in [a, b]$ define, for $V = C[a, b]$

$$E_{t_1} : V \to \mathbb{R}, \quad (E_{t_1}, f) = f(t_1) \tag{3}$$

It is easy to see that E_{t_1} is linear.

One could think of E_{t_1} as an integral with the weight the Dirac's delta function $\delta(t - t_1)$. Except, δ is not a function! It is a distribution.

Linear combinations of evaluation functions: if $t_1, \ldots, t_n \in [a, b]$ and $c_1, \ldots, c_n \in F$ then $c_1 E_{t_1} + \ldots + c_n E_{t_n}$ defined by

$$(c_1 E_{t_1} + \ldots + c_n E_{t_n}, f) = c_1 f(t_1) + \ldots + c_n f(t_n)$$

is also a linear functional on $C[a, b]$.

3. On the linear space $C^\infty(a, b)$ of infinitely many times differentiable functions on an interval (a, b), fix some $t_0 \in [a, b]$; then $(\boldsymbol{\phi}, f) = f^{(k)}(t_0)$ is a linear functional, and so is $(\boldsymbol{\phi}, f) = \sum_{k=0}^n c_k f^{(k)}(t_0)$.

1.2. **The dual space.** Given two linear functionals ϕ, ψ on V and $c \in F$ we can define their addition $\phi + \psi$ and scalar multiplication $c\phi$ by

$$(\boldsymbol{\phi} + \boldsymbol{\psi}, \mathbf{x}) = (\boldsymbol{\phi}, \mathbf{x}) + (\boldsymbol{\psi}, \mathbf{x}), \quad (c\boldsymbol{\phi}, \mathbf{x}) = c(\boldsymbol{\phi}, \mathbf{x})$$

and $\phi + \psi$, $c\phi$ are also linear functionals on V (check!).

Moreover, it can be easily verified that the set of linear functional on V form a vector space over F.

Definition 2. *The vector space V' of all the linear functionals on V is called the* **dual** *space of V.*

Often, knowledge of all functionals on a space yields complete information about the space. In finite dimensional cases, this is clear if we think of the component functionals of a vector.

The operation $(\boldsymbol{\phi}, \mathbf{x})$ is also called a *pairing* between V' and V.

Note that

$$(\boldsymbol{\phi}, c\mathbf{x} + d\mathbf{y}) = c(\boldsymbol{\phi}, \mathbf{x}) + d(\boldsymbol{\phi}, \mathbf{y})$$
$$(c\boldsymbol{\phi} + d\boldsymbol{\psi}, \mathbf{x}) = c(\boldsymbol{\phi}, \mathbf{x}) + d(\boldsymbol{\psi}, \mathbf{x})$$

More generally:

Definition 3. *Let U, V be vector spaces over the same scalar field F. A function $B : U \times V \to F$ is called a* **bilinear functional** *if $B(\mathbf{u}, \mathbf{v})$ is linear in $\mathbf{u}$ for each fixed $\mathbf{v}$, and linear in $\mathbf{v}$ for each fixed $\mathbf{u}$, in other words*

$$B(c_1\mathbf{u}_1 + c_2\mathbf{u}_2, \mathbf{v}) = c_1 B(\mathbf{u}_1, \mathbf{v}) + c_2 B(\mathbf{u}_2, \mathbf{v})$$

and

$$B(\mathbf{u}, d_1\mathbf{v}_1 + d_2\mathbf{v}_2) = d_1 B(\mathbf{u}, \mathbf{v}_1) + d_2 B(\mathbf{u}, \mathbf{v}_2)$$

for all $\mathbf{u}, \mathbf{u}_{1,2} \in U$ and all $\mathbf{v}, \mathbf{v}_{1,2} \in V$ and scalars $c_j, d_j \in F$.

The pairing between V' and V is a bilinear functional.

1.3. **Dual basis.** Given $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ a basis in V, its dual is one special associated basis for V': consider the linear functionals $\mathbf{v}'_1, \ldots, \mathbf{v}'_n$ defined by $(\mathbf{v}'_j, \mathbf{v}_i) = \delta_{i,j}$ for all $i, j = 1, \ldots, n$, where $\delta_{i,j}$ is the Kronecker symbol:

$$\delta_{i,j} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Note that the row vector representations of $\mathbf{v}'_1, \mathbf{v}'_2, \ldots, \mathbf{v}'_n$ in the basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ are

$$[1, 0, \ldots, 0],\ [0, 1, \ldots, 0], \ldots, [0, 0, \ldots, 1]$$

In particular, it is clear that $\mathbf{v}'_1, \mathbf{v}'_2, \ldots, \mathbf{v}'_n$ are linearly independent.

Consider any functional $\phi \in V'$ and its row vector representation (1) in the basis $\mathcal{B}$. Then

$$\boldsymbol{\phi} = \sum_{j=1}^{n} \phi_j \mathbf{v}'_j \tag{4}$$

In particular, $\mathbf{v}'_1, \mathbf{v}'_2, \ldots, \mathbf{v}'_n$ span V'.

We have shown:

Theorem 4. $\mathbf{v}'_1, \ldots, \mathbf{v}'_n$ *is a basis for V'. Therefore dim V'=dim V.*

Definition 5. $\mathcal{B}' = \{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$ *is called the* **basis dual** *to* $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.

Notes.

1. Applying elements of $\mathcal{B}'$ to vectors produces their coordinates in $\mathcal{B}$:

$$(\mathbf{v}'_k, \mathbf{x}) = x_k \tag{5}$$

Applying functionals to vectors in $\mathcal{B}$ produces their coordinates in $\mathcal{B}'$:

$$(\boldsymbol{\phi}, \mathbf{v}_j) = \phi_j$$

2. If $\mathbf{x} = \sum_j x_j \mathbf{v}_j$ is the representation in $\mathcal{B}$, and $\phi = \sum_k \phi_k \mathbf{v}'_k$, is the representation in $\mathcal{B}'$ then

$$(\boldsymbol{\phi}, \mathbf{x}) = \phi_1 x_1 + \ldots + \phi_n x_n = [\phi_1, \ldots, \phi_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

1.4. **Linear functionals as covectors; change of basis.** Suppose $\mathcal{B}_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V. Consider a linear functional $\phi : V \to F$ with the row vector representation (2) in this basis.

Let $\tilde{\mathcal{B}}_V = \{\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_n\}$ be another basis in V linked to $\mathcal{B}_V$ by the transition matrix S: $\tilde{\mathbf{v}}_j = \sum_i S_{ij} \mathbf{v}_i$.

Recall that the coordinates of vectors change with S^{-1}: if $\mathbf{x} = \sum_{j=1}^n x_j \mathbf{v}_j = \sum_{j=1}^n \tilde{x}_j \tilde{\mathbf{v}}_j$ then

$$\begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{bmatrix} = S^{-1} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

while matrix representations of linear transformations changes with S:

$$[\tilde{\phi}_1, \ldots, \tilde{\phi}_n] = [\phi_1, \ldots, \phi_n]\, S$$

(where $\tilde{\phi}_j = (\boldsymbol{\phi}, \tilde{\mathbf{v}}_j)$).

Functionals are covariant, while vectors are contravariant. This is one reason functionals are also called **covectors**. Sometimes, covectors are simply written as rows, acting on columns by the usual matrix multiplication.

Note: if ϕ is a functional and $\mathbf{x}$ is a vector, we can use their representations in different bases to calculate $(\boldsymbol{\phi}, \mathbf{x})$:

$$(\boldsymbol{\phi}, \mathbf{x}) = [\phi_1, \ldots, \phi_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [\tilde{\phi}_1, \ldots, \tilde{\phi}_n] \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{bmatrix}$$

1.4.1. *Change of the dual basis upon a change of coordinates.* To find the transition matrix from the basis dual to $\mathcal{B}_V$ to the one dual to $\tilde{\mathcal{B}}_V$ note that

$$(\tilde{\mathbf{v}}'_j, \mathbf{v}_k) = (\tilde{\mathbf{v}}'_j, \sum_l S^{-1}_{lk} \tilde{\mathbf{v}}_l) = \sum_l S^{-1}_{lk} (\tilde{\mathbf{v}}'_j, \tilde{\mathbf{v}}_l) = \sum_l S^{-1}_{lk} \delta_{jl} = S^{-1}_{jk}$$

$$= \sum_i S^{-1}_{ji} \delta_{ik} = \sum_i S^{-1}_{ji} (\mathbf{v}'_i, \mathbf{v}_k) = (\sum_i S^{-1}_{ji} \mathbf{v}'_i, \mathbf{v}_k)$$

for all k, hence $\tilde{\mathbf{v}}'_j = \sum_i S^{-1}_{ji} \mathbf{v}'_i$ therefore the transition matrix between the dual bases is $(S^{-1})^T$.

Exercise. How does this compare with the matrix change of variables $S^{-1}MS$? Why do we get something different?

1.5. **Functionals and hyperplanes.** Two-dimensional subspaces in $\mathbb{R}^3$ are planes going through the origin; they can be specified as the set of all vectors orthogonal to a given vector $\mathbf{n} \in \mathbb{R}^3$ as $\{\mathbf{x} \in \mathbb{R}^3 \,|\, \mathbf{n} \cdot \mathbf{x} = (\mathbf{n}^T, \mathbf{x}) = 0\}$. The row vector $\mathbf{n}^T$ is the transpose of $\mathbf{n}$.

The plane in $\mathbb{R}^3$ passing through a point $\mathbf{x}_0$ and orthogonal to $\mathbf{n} \in \mathbb{R}^3$ is $\{\mathbf{x} \in \mathbb{R}^3 \,|\, (\mathbf{n}^T, \mathbf{x} - \mathbf{x}_0) = 0\}$. Equivalently, this is $\{\mathbf{x} \in \mathbb{R}^3 \,|\, (\mathbf{n}^T, \mathbf{x}) = k\}$ where $k = (\mathbf{n}^T, \mathbf{x}_0)$. Also, equivalently, this is a translation of a two-dimensional subspace: $\mathbf{x}_0 + \{\mathbf{x} \in \mathbb{R}^3 \,|\, (\mathbf{n}^T, \mathbf{x}) = 0\}$.

This situation is generalized in arbitrary dimensions as follows.

Hyperspaces are subspaces on $\mathbb{R}^n$ of dimension $n-1$ (they are said to have *codimension* 1). Hyperspaces H of a vector space V are maximal subspaces: if U is a subspace so that $H \subset U \subset V$ then $U = H$ or $U = V$.

Hyperplanes are translations of hyperspaces.

Theorem 6. *Let V be a vector space.*

a) Every hyperspace in V is the null space of a linear functional.

b) Every hyperplane in V is the level set of a linear functional. In other words, a hyperplane is the set $H_{\phi,k} = \{\mathbf{x} \in V \,|\, (\boldsymbol{\phi}, \mathbf{x}) = k\}$ where $\phi \in V'$ is a nonzero *linear functional and $k \in F$.*

We may prove these facts a bit later.

1.6. **Application: Spline interpolation of sampled data (*Based on Prof. Gerlach's notes*).** Suppose we measure some data: at $n+1$ data points $s_0, s_1, \ldots, s_n$, assumed *distinct*, we measure the quantities $y_0, y_1, \ldots, y_n$. The question we want to address is to fit them on a curve, namely to find a function $\psi(t)$ so that

$$\psi(x_j) = y_j \text{ for all } j = 0, 1, \ldots, n \tag{6}$$

Of course, there are infinitely many such functions ψ, and to have a definite (unique) answer, we need to restrict our search to special classes of functions. The choice of the type of functions used for interpolation is done taking into account expected error, calculation time etc.

1.6.1. *Polynomial interpolation.* Polynomials in $\mathcal{P}_n$, of degree at most n, depend on $n+1$ parameters, so it is reasonable to expect one can find a unique polynomial $\psi \in \mathcal{P}_n$ satisfying (6). A polynomial of degree n passing through the given points is called *spline polynomial* of degree n. Before computers, elastic strips were used to produce the interpolation. Look up on Wikipedia "flat spline".

To find the spline polynomial, we will seek a general formula for a polynomial taking the value p_j at $x = x_j$.

A general formula is derived as follows. Consider the evaluation linear functionals (3) on $V = \mathcal{P}_n$: $E_{s_0}, \ldots, E_{s_n}$ which satisfy $(E_{s_j}, p) = p(s_j)$ for all $j = 0, \ldots, , n$ and all $p \in \mathcal{P}_n$.

The evaluation functional $E_{s_0}, \ldots, E_{s_n}$ are linearly independent: assuming a linear combination $\sum_k c_k E_{s_k} = 0$, then in particular, $(\sum_k c_k E_{s_k}, t^j) = 0$ for all $j = 0, 1, \ldots, n$ therefore $\sum_k c_k {s_k}^j = 0$ for all $j = 0, 1, \ldots, n$. This is a homogeneous system of $n+1$ equations with $n+1$ unknowns $c_0, \ldots, c_n$ whose determinant is a Vandermonde determinant, equal to $\Pi_{i>j}(s_i - s_j) \neq 0$, and therefore all $c_j = 0$. Since $E_{s_0}, \ldots, E_{s_n}$ are an $n+1$ set of independent elements in the $n+1$ dimensional linear space $\mathcal{P}'_n$, then they form a basis.

Let us find the basis $Q_0, \ldots, Q_n$ of $\mathcal{P}_n$ for which $E_{s_0}, \ldots, E_{s_n}$ is the dual basis. It means $(E_{s_j}, Q_k) = \delta_{jk}$. Once we accomplish that, then any polynomial $p \in \mathcal{P}_n$ can be written as $p = \sum_k \alpha_k Q_k$, on the one hand. On the other hand, $(E_{s_j}, p) = (E_{s_j}, \sum_k \alpha_k Q_k) = \alpha_k$ and thus $p = \sum_k (E_{s_k}, p) Q_k$ (by (5)), therefore

$$p = \sum_k (E_{s_k}, p) Q_k = \sum_{k=0}^{n} p(s_k) Q_k$$

which gives any polynomial $p \in \mathcal{P}_n$ based on its sample values $p(s_0), \ldots, p(s_n)$.

To find the basis $Q_0, \ldots, Q_n$, note that the polynomial Q_k must satisfy $(E_{s_j}, Q_k) = Q_k(s_j) = 0$ for all $j \neq k$, therefore $Q_k = c_k(t - s_0)(t - s_1) \ldots (t - s_{j-1})(t - s_{j+1}) \ldots (t - s_n)$ with c_k determined so that $Q_k(s_k) = 1$, therefore

$$Q_k = \frac{(t - s_0)(t - s_1) \ldots (t - s_{j-1})(t - s_{j+1}) \ldots (t - s_n)}{(s_k - s_0)(s_k - s_1) \ldots (s_k - s_{j-1})(s_k - s_{j+1}) \ldots (s_k - s_n)}$$

In conclusion, there is a unique polynomial $\psi \in \mathcal{P}_n$ satisfying (6), namely

$$\psi(t) = \sum_{k=0}^{n} y_k Q_k(t) \tag{7}$$

which is called the *Lagrange formula for polynomial interpolation.*

Note: the fact that there exists a unique interpolating function in a certain linear space is based on the fact that the evaluation functionals form a basis for that linear space.

1.6.2. *Linear interpolation.* A simpler interpolating function can be obtained by joining each consecutive sample points (s_j, y_j) and (s_{j+1}, y_{j+1}) by line segments; the result is an interpolating function which is continuous, and piecewise linear (but not differentiable).

Of course, it is easy to write formulas, and such an interpolating function satisfying (6), is

$$\psi(t) = y_j + \frac{y_{j+1} - y_j}{s_{j+1} - s_j} (t - s_j) \text{ for all } t \in [s_j, s_{j+1}]$$

Let us see how we can write ψ as a linear combination of basic functions, an analogue of (7) from the polynomial interpolation case.

Given $s_0 < s_1 < \ldots < s_n$ we look for an interpolating function in the space

$$V = \left\{ f : [s_0, s_n] \to \mathbb{R} \,\middle|\, f(t) = \text{ a line segment for } t \in [s_j, s_{j+1}],\ j = 0, \ldots, n \right\}$$

This is a linear space, of dimension $n + 1$ (you may wish to check!).

As in §1.6.1, we find the basis $F_0, \ldots, F_n$ of V for which $E_{s_0}, \ldots, E_{s_n}$ is the dual basis. (It is quite clear that these evaluations form a basis for V since any function is completely determined by its values at the sample points.) The piecewise linear function F_k must satisfy $(E_{s_j}, F_k) = F_k(s_j) = 0$ for all $j \neq k$, and $F_k(s_k) = 1$. Therefore F_k is a "tent" function which is zero on $[s_0, s_{j-1}]$ and on $[s_{j+1}, s_n]$ and whose graph joins by a segment the points $(s_{j-1}, 0)$ and $(s_j, 1)$ and by a segment the points $(s_j, 1)$ and $(s_{j+1}, 0)$.

Then

$$\psi(t) = \sum_{k=0}^{n} y_k F_k(t)$$

is the interpolating function in V for the data (6).

1.6.3. *Band-limited interpolation.* Another type of interpolating function is a superposition of oscillations. We assume here the samples are taken over an interval which we take to be 2π (for simplicity of formulas).

We look for interpolating functions in the linear space spanned by the linearly independent functions 1, $\sin(kt)$, $\cos(kt)$, $k = 1, \ldots, N$:

$$\mathcal{B}_N = \left\{ f \mid f(x) = \sum_{k=0}^{N} a_k \cos(kt) + \sum_{k=1}^{N} b_k \sin(kt),\ a_k, b_k \in \mathbb{R} \right\}$$

Since this an $2N + 1$ dimensional space, we need an equal number of samples: $s_0, s_1, \ldots, s_{2N} \in [0, 2\pi]$.

As in the examples before, we look for a basis $S_k,\ k = 0, \ldots, 2N$ of $\mathcal{B}_N$ so that the evaluation functionals E_{s_k} from its dual basis.[1] The function S_k must satisfy $S_k(x_j) = 0$ for all $j \neq k$. By analogy with the Lagrange formula for polynomial interpolation a first attempt may be to think of S_k as a scalar multiple of the product of all $\sin(t - s_j)$ with $j \neq k$; however this does not work because this product belongs to $\mathcal{B}_{2N}$ rater than $\mathcal{B}_N$. What works however is to take S_k as a scalar multiple of the product of all $\sin \frac{1}{2}(t - s_j)$ with $j \neq k$.

Indeed, first note that this product belongs to $\mathcal{B}_N$. There is an even number of factors in the product, and a product of any two of them is $\sin \frac{1}{2}(t - s_j) \sin \frac{1}{2}(t - s_i) = \cos \frac{1}{2}(s_i - s_j) - \cos(t - \frac{s_i + s_j}{2}) \in \mathcal{B}_N$. Thus S_k is the product of N terms of the form $a_k + b_k \sin t + c_k \cos t$ which is known to have the form of functions in $\mathcal{B}_N$ (e.g. $2 \sin x \cos x = \sin(2x)$, $2 \cos^2 x =$

[1] In this example we will not prove that the evaluations form a basis, rather attempt to produce the basis for $\mathcal{B}_N$; we will succeed, proving that the evaluations were a basis indeed.

$1+\cos(2x)$, $4\cos^3 x = \cos(3x)+\cos x$ etc.). Normalizing to ensure that $S_k(x_k)=1$ we obtain the formula

$$S_k(t) = \prod_{j=0,\, j\neq k}^{2N} \frac{\sin\frac{1}{2}(t-s_j)}{\sin\frac{1}{2}(s_k-s_j)}$$

and the interpolating function is

$$\psi(t) = \sum_{k=0}^{2N} y_k S_k(t)$$

1.7. **The bidual.** The dual V' is a vector space, therefore it has a dual V'', called the bidual of V. It turns out that the bidual can be naturally identified with V by the pairing $(\boldsymbol{\phi},\mathbf{x})$, which can be interpreted either as ϕ acting on $\mathbf{x}$, or as $\mathbf{x}$ acting on ϕ. First we show the following result, which is sometimes phrased as "linear functionals separate points".

Lemma 7. *If $\mathbf{x}\in V$ is so that $(\boldsymbol{\phi},\mathbf{x})=0$ for all $\phi\in V'$, then $\mathbf{x}=\mathbf{0}$.*

To prove the Lemma, let $\mathbf{v}_1,\ldots,\mathbf{v}_n$ of a basis of V; then $\mathbf{x}=\sum_j x_j\mathbf{v}_j$ and we have in particular, $(\mathbf{v}'_k,\mathbf{x})=x_k=0$ for all k, therefore $\mathbf{x}=\mathbf{0}$.

Theorem 8. *The bidual V'' of a finite dimensional space V is isomorphic to V.*

Why: formalizing the intuitive argument above, let

$$V'' = \{L : V' \to F \,|, L \text{ linear}\}$$

and define $T: V\to V''$ as follows. For each $\mathbf{x}\in V$ let $T\mathbf{x}\in V''$ be defined by $(T\mathbf{x})\phi = (\boldsymbol{\phi},\mathbf{x})$. Clearly T is linear. To show T is 1-to-1, assume that there is an $\mathbf{x}\in V$ such that $T\mathbf{x}=\mathbf{0}$. This means that $(T\mathbf{x})\phi=(\boldsymbol{\phi},\mathbf{x})=0$ for all $\phi\in V'$. It follows that $\mathbf{x}=\mathbf{0}$. Since T is injective and since $\dim V = \dim V' = \dim V''$, then T is also onto, hence it is an isomorphism. $\Box$.

Note that the construction of the isomorphism of Theorem 8 relies on no particular basis: this is called a *natural isomorphism.*

1.8. **Orthogonality.** The construction of hyperplanes suggests that the dual of a vector space can be used to define an orthogonality-like relation. For example,

Definition 9. *If U is a subspace of a vector space V the* annihilator *of U is the set*

$$U^\perp = \{\phi\in V' \,|\, (\boldsymbol{\phi},\mathbf{u})=0,\ \textit{for all } \mathbf{u}\in U\}$$

It can be checked that the annihilator $U^\perp$ is a subspace of V'. In absence of an inner product, the annihilator works quite well.

1.9. **The transpose.**

Warning: change of notation. Consistent notations are useful for helping the thought process focus on the essential features rather than mere names, and these notes have tried to use consistent notations. However...

Up to now we used consistently the letter T to denote linear transformations. At this point we are going to encounter the transpose of matrices, denoted by a superscript T. To avoid collisions of notations, from now on linear transformations will be denoted by the letter L.

Definition 10. *If $L : U \to V$ is a linear transformation between two vector spaces U, V over the scalar field F then the* dual transformation, *or the* **transpose transformation** *is the linear transformation $L' : V' \to U'$ defined by*

$$(L'\boldsymbol{\phi}, \mathbf{x}) = (\boldsymbol{\phi}, L\mathbf{x}) \quad \textit{for all } \mathbf{x} \in U, \phi \in V' \tag{8}$$

As the words suggest:

Theorem 11. *If M is the matrix of $L : U \to V$ in the bases $\mathcal{B}_U$, $\mathcal{B}_V$ then M^T is the matrix of L' in the dual bases $\mathcal{B}'_V$, $\mathcal{B}'_U$.*

Formal proof. Since M is the matrix of L we have $L\mathbf{u}_k = \sum_i M_{ik}\mathbf{v}_i$. To calculate $L'\mathbf{v}'_j$ we apply it to the vectors $\mathbf{x}$:

$$(L'\mathbf{v}'_j, \mathbf{x}) = (L'\mathbf{v}'_j, \sum_k x_k\mathbf{u}_k) = \sum_k x_k(L'\mathbf{v}'_j, \mathbf{u}_k) = \sum_k x_k(\mathbf{v}'_j, L\mathbf{u}_k)$$

$$= \sum_k x_k(\mathbf{v}'_j, \sum_i M_{ik}\mathbf{v}_i) = \sum_k\sum_i x_k M_{ik}(\mathbf{v}'_j, \mathbf{v}_i) = \sum_k\sum_i x_k M_{ik}\delta_{ij}$$

$$= \sum_k x_k M_{jk} = \sum_k M_{jk}(\mathbf{u}'_k, \mathbf{x}) = (\sum_k M_{jk}\mathbf{u}'_k, \mathbf{x}) \text{ for all } \mathbf{x}$$

which shows that $L'\mathbf{v}_j = \sum_k M_{jk}\mathbf{u}'_k$, which proves the theorem. □

1.9.1. *The four fundamental spaces of a matrix.* Consider an $m \times n$ matrix M with entries in F ($F = \mathbb{R}$ or $\mathbb{C}$) and its transpose M^T. To rewrite (8) in matrix notation it is sometimes preferred to replace the row vectors ϕ by transposes of (usual) column vectors, $\mathbf{y}^T$. Relation (8) then reads $(M^T\mathbf{y})^T\mathbf{x} = \mathbf{y}^T M\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$ (or $\mathbb{C}^n$) and all $\mathbf{y} \in \mathbb{R}^m$ (respectively $\mathbb{C}^M$), a relation which is clear bases on the basic property that $(AB)^T = B^T A^T$ for any two matrices A, B for which the multiplication makes sense.

Recall that we defined:

$\mathcal{R}(M) = \{\mathbf{y} \in F^m \,|\, M\mathbf{x} = \mathbf{y}$ for some $\mathbf{x} \in F^m\}$ **the column space** of M

$\mathcal{N}(M) = \{\mathbf{x} \in F^n \,|\, M\mathbf{x} = \mathbf{0}\}$ **the null space** of M

Since the rows of M^T are the columns of M, then clearly the row space of M^T equals the column space of M, and column space of M^T equals the row space of M:

$\mathcal{R}(M^T) =$ **the row space** of M.

Define
$\mathcal{N}(M^T) = \{\mathbf{y} \in\ F^m \,|\, M^T\mathbf{y} = \mathbf{0}\}$ **the left null space** of M

Note that $M^T\mathbf{y} = \mathbf{0}$ is equivalent to $\mathbf{y}^T M = \mathbf{0}$, justifying the name of *left null space*.

Recall that $\dim \mathcal{R}(M) + \dim \mathcal{N}(M) = n$ and using this for M^T, we obtain $\dim \mathcal{R}(M^T) + \dim \mathcal{N}(M^T) = m$.

Recall that $\dim \mathcal{R}(M) = rank(M)$ and the rank of a matrix is the order of the largest nonzero minor; then $rank(M) = rank(M^T)$.

Summarizing:

Theorem 12. The fundamental theorem of linear algebra

Let M be an $m \times n$ matrix. Ler $r = rank(M)$. Then

$$\begin{aligned}
&\dim \mathcal{R}(M) = r\\
&\dim \mathcal{R}(M^T) = r\\
&\dim \mathcal{N}(M) = n - r\\
&\dim \mathcal{N}(M^T) = m - r
\end{aligned}$$

Also

$$\begin{aligned}
\mathcal{N}(M) &= \ \mathcal{R}(M^T)^\perp\\
\mathcal{N}(M^T) &= \ \mathcal{R}(M)^\perp
\end{aligned}$$

where $\perp$ signifies the annihilator of the set.

Only the second to last line needs a proof. Here it is: consider first $\mathbf{x} \in \mathcal{N}(M)$ and show that $\mathbf{x}$ belongs to the annihilator of $\mathcal{R}(M^T)$, in other words, that $\mathbf{x}^T(M^T\mathbf{y}) = 0$ for all $\mathbf{y} \in F^m$, which is obvious since $\mathbf{x}^T(M^T\mathbf{y}) = (M\mathbf{x})^T\mathbf{y}$. Conversely, taking $\mathbf{x} \in \mathcal{R}(M^T)^\perp$, this means that $0 = \mathbf{x}^T(M^T\mathbf{y}) = (M\mathbf{x})^T\mathbf{y}$ for all $\mathbf{y} \in F^m$, which means that $M\mathbf{x} = \mathbf{0}$, hence $\mathbf{x} \in \mathcal{N}(M)$. $\square$

1.10. **Eigenvalues of the transpose.** As a corollary of Theorem 11, L and L' have the same eigenvalues. The following theorem shows that eigenvectors corresponding to different eigenvalues are orthogonal-like:

Theorem 13. *Let V be a finite dimensional space over $\mathbb{C}$, and $L : V \to V$ be a linear transformation. Let L' be its transpose, $L' : V' \to L'$.*

Let $\lambda_1 \neq \lambda_2$ be two eigenvalues of L, with eigenvectors $\mathbf{v}_1$ of L and ϕ_2 of L':

$$L(\mathbf{v}_1) = \lambda_1 \mathbf{v}_1, \quad L'(\phi_2) = \lambda_2 \phi_2$$

Then

$$(\boldsymbol{\phi}_2, \mathbf{v}_1) = 0$$

In particular, if $\lambda_1 \neq \lambda_2$ are eigenvalues of the $n \times n$ matrix M (and therefore of M^T as well), and if $M\mathbf{v} = \lambda_1 \mathbf{v}$ and $M^T\mathbf{y} = \lambda_2 \mathbf{y}$ then

$$(\mathbf{y}, \mathbf{v}) := y_1 v_1 + \ldots + y_n v_n = 0$$

Proof.

By the definition of the transpose transformation (see (8)) we must have

$$(L'\phi_2, \mathbf{v}_1) = (\boldsymbol{\phi}_2, L\mathbf{v}_1)$$

therefore $(\lambda_2 \phi_2, \mathbf{v}_1) = (\boldsymbol{\phi}_2, \lambda_1 \mathbf{v}_1)$ so $\lambda_2(\boldsymbol{\phi}_2, \mathbf{v}_1) = \lambda_1(\boldsymbol{\phi}_2, \mathbf{v}_1)$ and sinve $\lambda_1 \neq \lambda_2$ the theorem follows. $\square$