Vector Spaces

Throughout this course, the base field $\mathbb{F}$ of scalars will be $\mathbb{R}$ or $\mathbb{C}$. Recall that a vector space is a nonempty set V on which one defines the operations of addition (for $v,\ w \in V$, $v + w \in V$) and scalar multiplication (for $\alpha \in \mathbb{F}$ and $v \in V$, $\alpha v \in V$) which satisfy the following axioms: for every $x, y, z \in V$ and $\lambda, \mu \in \mathbb{F}$

1. $(x + y) + z = x + (y + z)$,
2. $x + y = y + x$,
3. $\exists\, 0 \in V$ such that $x + 0 = x$,
4. $\forall\, x \in V$ there exists $z \in V$ such that $x + z = 0$ (written $z = -x$),
5. $\lambda x = x\lambda$,
6. $\lambda(x + y) = \lambda x + \lambda y$,
7. $(\lambda + \mu)x = \lambda x + \mu x$,
8. $\lambda(\mu x) = (\lambda\mu)x$, and
9. $1x = x$.

A subset $W \subset V$ is a *subspace* if W is closed under addition and scalar multiplication, so W inherits a vector space structure of its own.

Examples:

(1) $\mathbb{F}^n = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : \text{ each } x_j \in \mathbb{F} \right\}, n \geq 1$

(2) $\{0\}$

(3) $\mathbb{F}^\infty = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} : \text{each } x_j \in \mathbb{F} \right\}$

(4) $\ell^1(\mathbb{F}) \subset \mathbb{F}^\infty$, where $\ell^1(\mathbb{F}) = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} : \sum_{j=1}^{\infty} |x_j| < \infty \right\}$

$\ell^\infty(\mathbb{F}) \subset \mathbb{F}^\infty$, where $\ell^\infty(\mathbb{F}) = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} : \sup_j |x_j| < \infty \right\}$

$\ell^1(\mathbb{F})$ and $\ell^\infty(\mathbb{F})$ are clearly subspaces of $\mathbb{F}^\infty$.

Let $0 < p < \infty$, and define $\ell^p(\mathbb{F}) = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} : \sum_{j=1}^{\infty} |x_j|^p < \infty \right\}$.

Since

$$\begin{aligned} |x+y|^p &\leq (|x| + |y|)^p \leq (2\max(|x|,|y|))^p \\ &= 2^p \max(|x|^p, |y|^p) \leq 2^p(|x|^p + |y|^p), \end{aligned}$$

it follows that $\ell^p(\mathbb{F})$ is a subspace of $\mathbb{F}^\infty$.

Exercise: Show that $\ell^p(\mathbb{F}) \subsetneqq \ell^q(\mathbb{F})$ if $0 < p < q \leq \infty$.

(5) Let X be a nonempty set; then the set of all functions $f : X \to \mathbb{F}$ has a natural structure as a vector space over $\mathbb{F}$: define $f_1 + f_2$ by $(f_1 + f_2)(x) = f_1(x) + f_2(x)$, and define αf by $(\alpha f)(x) = \alpha f(x)$.

(6) For a metric space X, let $C(X, \mathbb{F})$ be the set of all continuous $\mathbb{F}$-valued functions on X. $C(X, \mathbb{F})$ is a subspace of the vector space defined in (5). Define $C_b(X, \mathbb{F}) \subset C(X, \mathbb{F})$ to be the subspace of all bounded continuous functions $f : X \to \mathbb{F}$, and let $C^k(X, F) \subset C(X, \mathbb{F})$ to be the subspace of all k times continuously differentiable functions $f : X \to \mathbb{F}$. In the case where $X = \mathbb{F}$ we simplify this notations to $C(\mathbb{F})$, $C_b(\mathbb{F})$, and $C^k(\mathbb{F})$, respectively.

(7) Define $\mathcal{P}(\mathbb{F}) \subset C(\mathbb{R}, \mathbb{F})$ to be the space of all $\mathbb{F}$-valued polynomials on $\mathbb{R}$:

$$\mathcal{P}(\mathbb{F}) = \{a_0 + a_1 x + \cdots + a_m x^m : m \geq 0, \text{ each } a_j \in \mathbb{F}\}.$$

Each $p \in \mathcal{P}(\mathbb{F})$ is viewed as a function $p : \mathbb{R} \to \mathbb{F}$ given by $p(x) = a_0 + a_1 x + \cdots + a_m x^m$.

(8) Define $\mathcal{P}_n(\mathbb{F}) \subset \mathcal{P}(\mathbb{F})$ to be the subspace of all polynomials of degree $\leq n$.

(9) Let $V = \{u \in C^2(\mathbb{F}) : u'' + u = 0\}$. It is easy to check directly from the definition that V is a subspace of $C^2(\mathbb{F})$. For $\mathbb{F} = \mathbb{C}$, one knows that

$$V = \{a_1 \cos x + a_2 \sin x : a_1, a_2 \in \mathbb{C}\} = \{b_1 e^{ix} + b_2 e^{-ix} : b_1, b_2 \in \mathbb{C}\},$$

from which it is also clear that V is a vector space.

More generally, if $L(u) = u^{(m)} + a_{m-1} u^{(m-1)} + \cdots + a_1 u' + a_0 u$ is an m^{th} order linear constant-coefficient differential operator, then $V = \{u \in C^m(\mathbb{F}) : L(u) = 0\}$ is a vector space. V can be explicitly described as the set of all linear combinations of certain functions of the form $x^j e^{rx}$ where $j \geq 0$ and r is a root of the characteristic polynomial $r^m + a_{m-1} r^{m-1} + \cdots + a_1 r + a_0 = 0$. For details, see Chapter 3 of Birkhoff & Rota.

Convention: Throughout this course, if the field $\mathbb{F}$ is not specified, it is assumed to be $\mathbb{C}$.

Linear Independence, Span, Basis

Let V be a vector space. A *linear combination* of the vectors $v_1, \ldots, v_m \in V$ is a vector $v \in V$ of the form $v = \alpha_1 v_1 + \cdots + \alpha_m v_m$ where each $\alpha_j \in \mathbb{F}$. Let $S \subset V$ be a subset of V. S is called *linearly independent* if for every finite subset $\{v_1, \ldots, v_m\}$ of S, the linear combination $\sum_{i=1}^m \alpha_i v_i = 0$ iff $\alpha_1 = \cdots = \alpha_m = 0$. Otherwise, S is called *linearly dependent*. Define the *span of* S (denoted $\text{Span}(S)$) to be the set of all linear combinations of all finite subsets of S. (Note: a linear combination is by definition a *finite* sum.) If $S = \emptyset$, set $\text{Span}(S) = \{0\}$. S is said to be a *basis* of V if S is linearly independent and $\text{Span}(S) = V$.

Facts: (a) Every vector space has a basis; in fact if S is any linearly independent set in V, then there is a basis of V containing S. The proof of this in infinite dimensions uses Zorn's lemma and is nonconstructive. Such a basis in infinite dimensions is called a *Hamel basis.* Typically it is impossible to identify a Hamel basis explicitly, and they are of little use. There are other sorts of "bases" in infinite dimensions defined using topological considerations which are very useful and which we will consider later.

(b) Any two bases of the same vector space V can be put into $1-1$ correspondence. Define the *dimension of* V (denoted $\dim V$) $\in \{0, 1, 2, \ldots\} \cup \{\infty\}$ to be the number of elements in a basis of V. The vectors $e_1, \ldots, e_n$, where

$$e_j = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow j^{\text{th}} \text{ entry},$$

form the *standard basis* of $\mathbb{F}^n$, and $\dim \mathbb{F}^n = n$.

Remark. Any $\mathbb{C}$-vector-space V may be regarded as an $\mathbb{R}$-vector-space by restriction of the scalar multiplication. It is easily checked that if V is finite-dimensional with basis $\{v_1, \ldots, v_n\}$ over $\mathbb{C}$, then $\{v_1, \ldots, v_n, iv_1, \ldots, iv_n\}$ is a basis for V over $\mathbb{R}$. In particular, $\dim_{\mathbb{R}} V = 2 \dim_{\mathbb{C}} V$.

The vectors $e_1, e_2, \ldots, \in \mathbb{F}^\infty$ are linearly independent. However, $\text{Span}\{e_1, e_2, \ldots\}$ is the proper subset of $\mathbb{F}^\infty$ consisting of all vectors with only finitely many nonzero components, so $\{e_1, e_2, \ldots\}$ is not a basis of $\mathbb{F}^\infty$. But $\{x^m : m \in \{0, 1, 2, \ldots\}\}$ *is* a basis of $\mathcal{P}$.

Now let V be a finite-dimensional vector space, and $\{v_1, \ldots, v_n\}$ be a basis for V. Any $v \in V$ can be written uniquely as $v = \sum_{i=1}^n x_i v_i$ for some $x_i \in \mathbb{F}$. So we can define a map from V into $\mathbb{F}^n$ by $v \mapsto \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. The x_i's are called the *coordinates* of v with respect to the basis $\{v_1, \ldots, v_n\}$. This coordinate map clearly preserves the vector space operations and is bijective, so it is an isomorphism of V with $\mathbb{F}^n$ in the following sense.

Definition. Let V, W be vector spaces. A map $L : V \to W$ is a *linear transformation* if

$$(\forall\, v_1, v_2 \in V)(\forall\, \alpha_1, \alpha_2 \in \mathbb{F})\ \ L(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 L(v_1) + \alpha_2 L(v_2).$$

If in addition L is bijective, then L is called a (vector space) *isomorphism.*

Even though every finite-dimensional vector space V is isomorphic to $\mathbb{F}^n$ (where $n = \dim V$), the isomorphism depends on the choice of basis. Many properties of V are independent of the basis (e.g. $\dim V$). We could try to avoid bases, but it is very useful to use coordinate systems. So we need to understand how coordinates change when the basis is changed.

Change of Basis

Let V be a finite dimensional vector space. Let $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_n\}$ be two bases for V. For $v \in V$, let $x = (x_1, \ldots, x_n)^T$ and $y = (y_1, \ldots, y_n)^T$ denote the vectors of coordinates of v with respect to the bases $\mathcal{B}_1 = \{v_1, \ldots, v_n\}$ and $\mathcal{B}_2 = \{w_1, \ldots, w_n\}$, respectively. So $v = \sum_{i=1}^n x_i v_i = \sum_{j=1}^n v_j w_j$. Express each w_j in terms of $\{v_1, \ldots, v_n\}$: $w_j = \sum_{i=1}^n a_{ij} v_i$ $(a_{ij} \in \mathbb{F})$. Let $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \in \mathbb{F}^{n \times n}$. Then

$$\sum_{i=1}^n x_i v_i = v = \sum_{j=1}^n y_j w_j = \sum_{i=1}^n \left(\sum_{j=1}^n a_{ij} y_j \right) v_i,$$

so $x_i = \sum_{j=1}^n a_{ij} y_j$, i.e. $x = Ay$. The matrix A is called the *change of basis* matrix.

Notation: Horn-Johnson uses $M_{m,n}(\mathbb{F})$ to denote $\mathbb{F}^{m \times n}$ = set of $m \times n$ matrices with entries from $\mathbb{F}$. H-J writes $[v]_{\mathcal{B}_1}$ for x, $[v]_{\mathcal{B}_2}$ for y, and ${}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}$ for A, so $x = Ay$ becomes $[v]_{\mathcal{B}_1} = {}_{\mathcal{B}_1}[I]_{\mathcal{B}_2}[v]_{\mathcal{B}_2}$.

Similarly, we can express each v_j in terms of $\{w_1, \ldots, w_n\}$: $v_j = \sum_{i=1}^n b_{ij} w_i$ $(b_{ij} \in \mathbb{F})$. Let $B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \in \mathbb{F}^{n \times n}$. Then $y = Bx$. We obtain that A and B are invertible and $B = A^{-1}$.

Formal matrix notation: Write the basis vectors $(v_1 \cdots v_n)$ and $(w_1 \cdots w_n)$ formally in rows. Then the equations $w_j = \sum_{i=1}^n a_{ij} v_i$ become the formal matrix equation $(w_1 \cdots w_n) = (v_1 \cdots v_n)A$ using the usual matrix multiplication rules. In general, $(v_1 \cdots v_n)$ and $(w_1 \cdots w_n)$ are not matrices (although in the special case where each v_j and w_j is a column vector in $\mathbb{F}^n$, we have $W = VA$ where $V, W \in \mathbb{F}^{n \times n}$ are the matrices whose columns are the v_j's and the w_j's, respectively). We also have the formal matrix equations $v = (v_1 \cdots v_n)x$ and $v = (w_1 \cdots w_n)y$, so

$$(v_1 \cdots v_n)x = (w_1 \cdots w_n)y = (v_1 \cdots v_n)Ay,$$

which gives us $x = Ay$ as before.

Remark. We can read the matrix equation $W = VA$ as saying the j^{th} col. of W is the linear combination of the cols. of V whose coefficients are in the j^{th} col. of A.

Constructing New Vector Spaces from Given Ones

(1) The intersection of any family of subspaces of V is again a subspace: let $\{W_\gamma : \gamma \in G\}$ be a family of subspaces of V (where G is an index set); then $\bigcap_{\gamma\in G} W_\gamma$ is a subspace of V.

(2) *Sums of subspaces:* If W_1, W_2 are subspaces of V, then

$$W_1 + W_2 = \{w_1 + w_2 : w_1 \in W_1, w_2 \in W_2\}$$

is also a subspace, and $\dim(W_1 + W_2) + \dim(W_1 \cap W_2) = \dim W_1 + \dim W_2$. We say that the sum $W_1 + W_2$ is *direct* if $W_1 \cap W_2 = \{0\}$ (equivalent: for each $v \in W_1 + W_2$, there are unique $w_1 \in W_1$ and $w_2 \in W_2$ for which $v = w_1 + w_2$), and we write $W_1 \oplus W_2$ for $W_1 + W_2$. More generally, if $W_1, \ldots, W_n$ are subspaces of V, then $W_1 + \cdots + W_2 = \{w_1 + \cdots + w_2 : w_j \in W_j, 1 \leq j \leq n\}$ is a subspace. We say that the sum is *direct* if whenever $w_j \in W_j$ and $\sum_{j=1}^n w_j = 0$, then each $w_j = 0$, and we write $W_1 \oplus \cdots \oplus W_n$. Even more generally, if $\{W_\gamma : \gamma \in G\}$ is a family of subspaces of V, define $\sum_{\gamma\in G} W_\gamma = \text{span}\left(\bigcup_{\gamma\in G} W_\gamma\right)$. We say that the sum is direct if for each finite subset G' of G, whenever $w_\gamma \in W_\gamma$ for $\gamma \in G'$ and $\sum_{\gamma\in G'} w_\gamma = 0$, then each $w_\gamma = 0$ for $\gamma \in G'$ (equivalently: for each $\beta \in G$, $W_\beta \cap \left(\sum_{\gamma\in G, \gamma\neq\beta} W_\gamma\right) = \{0\}$).

(3) *Direct Products:* Let $\{V_\gamma : \gamma \in G\}$ be a family of vector spaces over $\mathbb{F}$. Define $V = \bigtimes_{\gamma\in G} V_\gamma$ to be the set of all functions $v : G \to \bigcup_{\gamma\in G} V_\gamma$ for which $(\forall\, \gamma \in G)\ v(\gamma) \in V_\gamma$. We write v_γ for $v(\gamma)$, and we write $v = (v_\gamma)_{\gamma\in G}$, or just $v = (v_\gamma)$. Define $v + w = (v_\gamma + w_\gamma)$ and $\alpha v = (\alpha v_\gamma)$. Then V is a vector space over $\mathbb{F}$. (Example: $G = \mathbb{N} = \{1, 2, \ldots\}$, each $V_n = \mathbb{F}$. Then $\bigtimes_{n\geq 1} V_n = \mathbb{F}^\infty$.)

(4) (*External*) *Direct Sums:* Let $\{V_\gamma : \gamma \in G\}$ be a family of vector spaces over $\mathbb{F}$. Define $\bigoplus_{\gamma\in G} V_\gamma$ to be the subspace of $\bigtimes_{\gamma\in G} V_\gamma$ consisting of those v for which $v_\gamma = 0$ except for finitely many $\gamma \in G$. (Example: For $n = 0, 1, 2, \ldots$ let $V_n = \text{span}(x^n)$ in $\mathcal{P}$. Then $\mathcal{P} = \bigoplus_{n\geq 0} V_n$.) Technicality: we should technically assume that for $\gamma \neq \beta$, $V_\gamma \cap V_\beta = \{0\}$; if not, rename the elements of each v_γ to make it true; the given definition avoids the technicality by using the direct product.

Facts: (a) If G is a finite index set, XV_γ and $\oplus V_\gamma$ are isomorphic. (b) If each W_γ is a subspace of V and the sum $\sum_{\gamma\in G} W_\gamma$ is direct, then it is naturally isomorphic to the external direct sum $\oplus W_\alpha$.

(5) *Quotients:* Let W be a subspace of V. Define on V the equivalence relation $v_1 \sim v_2$ if $v_1 - v_2 \in W$, and define the quotient to be the set V/W of equivalence classes. Let $v + W$ denote the equivalence class of v. Define a vector space structure on V/W by defining $\alpha_1(v_1 + W) + \alpha_2(v_2 + W) = (\alpha_1 v_1 + \alpha_2 v_2) + W$. Define the *codimension* of W in V by $\text{codim}(W) = \dim(V/W)$.

Dual Vector Spaces

Definition. Let V be a vector space. A *linear functional* on V is a function $f : V \to \mathbb{F}$ for which $f(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 f(v_1) + \alpha_2 f(v_2)$ for $v_1, v_2 \in V$, $\alpha_1, \alpha_2 \in \mathbb{F}$, i.e., a linear transformation from V to the 1-dimensional vector space $\mathbb{F}$.

Examples:

(1) Let $V = \mathbb{F}^n$, and let f be a linear functional on V. Set $f_i = f(e_i)$ for $1 \le i \le n$. Then for $x = (x_1, \ldots, x_n)^T = \sum_{i=1}^n x_i e_i \in \mathbb{F}^n$,

$$f(x) = \sum_{i=1}^n x_i f(e_i) = \sum_{i=1}^n f_i x_i = (f_1 f_2 \cdots f_n)x.$$

So the row vector $(f_1 \cdots f_n)$ is the matrix of f using the standard basis $\{e_1, \ldots, e_n\}$ on $\mathbb{F}^n$ (and the basis $\{1\}$ on $\mathbb{F}$).

(2) Let $V = \mathbb{F}^\infty$. Given an N and some $f_1, f_2, \ldots, f_N \in \mathbb{F}$, we could define $f(x) = \sum_{i=1}^N f_i x_i$ for $x \in \mathbb{F}^\infty$. However, not all linear functionals on $\mathbb{F}^\infty$ are of this form.

(3) Let $V = \ell^1(\mathbb{F})$. If $f \in \ell^\infty(\mathbb{F})$, then for $x \in \ell^1(\mathbb{F})$, $\sum_{i=1}^\infty |f_i x_i| \le (\sup |f_i|) \sum_{i=1}^\infty |x_i| < \infty$, so the sum $f(x) = \sum_{i=1}^\infty f_i x_i$ converges absolutely, defining a linear functional on $\ell^1(\mathbb{F})$. Similarly, if $V = \ell^\infty(\mathbb{F})$ and $f \in \ell^1(\mathbb{F})$, $f(x) = \sum_{i=1}^\infty f_i x_i$ defines a linear functional on $\ell^\infty(\mathbb{F})$.

(4) Let $X \subset \mathbb{R}^n$ and $x_0 \in X$. Then $f(u) = u(x_0)$ defines a linear functional on $C(X)$.

(5) If $-\infty < a < b < \infty$, $f(u) = \int_a^b u(x)dx$ defines a linear functional on $C([a, b])$.

Definition. If V is a finite-dimensional vector space, the dual space of V is the vector space V^* of all linear functionals on V, where $(\alpha_1 f_1 + \alpha_2 f_2)(v) = \alpha_1 f_1(v) + \alpha_2 f_2(v)$.

Remark. When V is infinite dimensional, the set of *all* linear functions is often called the *algebraic* dual space of V, as it depends only on the algebraic structure of V. We will be more interested in linear functionals related also to a topological structure on V. After introducing norms (which induce metrics on V), we will define V^* to be the vector space of all *continuous* linear functionals on V. (When V is finite dimensional, with any norm on V, every linear functional on V is continuous.)

Dual Basis in Finite Dimensions

Let V be a finite dimensional vector space, and let $\{v_1, \ldots, v_n\}$ be a basis for V. For $1 \le i \le n$, define linear functionals $f_i \in V^*$ by $f_i(v_j) = \delta_{ij}$ ($= 1$ for $i = j, = 0$ for $i \ne j$). Let $v \in V$, and let $x = (x_1, \ldots, x_n)^T$ be the vector of coordinates of v with respect to the basis $\{v_i, \ldots, v_n\}$, i.e., $v = \sum_{i=1}^n x_i v_i$. Then $f_i(v) = x_i$, i.e., f_i maps v into its coordinate x_i of v_i. Now if $f \in V^*$, let $a_i = f(v_i)$; then

$$f(v) = f(\sum x_i v_i) = \sum_{i=1}^n a_i x_i = \sum_{i=1}^n a_i f_i(v),$$

so $f = \sum_{i=1}^n a_i f_i$. This representation is unique (exercise), so $\{f_1, \ldots, f_n\}$ is a basis for V^*, called the *dual basis* to $\{v_1, \ldots, v_n\}$. We get $\dim V^* = \dim V$.

If we write the dual basis in a column $\begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$ and the coordinates $(a_1 \cdots a_n)$ of $f = \sum_{i=1}^n a_i f_i \in V^*$ in a row, then $f = (a_1 \cdots a_n) \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$. The defining equation of the dual basis is (matrix multiply, evaluate)

$$(*) \qquad \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} (v_1 \cdots v_n) = \begin{pmatrix} 1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & 1 \end{pmatrix} = I$$

Change of Basis and Dual Bases: Let $\{w_1, \ldots, w_n\}$ be another basis of V related to $\{v_1, \ldots, v_n\}$ by the change-of-basis matrix A, i.e., $(w_1 \cdots w_n) = (v_1 \cdots v_n)A$. Left-multiplying $(*)$ by A^{-1} and right-multiplying by A gives

$$A^{-1} \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} (w_1 \cdots w_n) = I.$$

Therefore,

$$\begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix} = A^{-1} \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} \quad \text{satisfies} \quad \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix} (w_1 \cdots w_n) = I$$

and so $\{g_1, \ldots, g_n\}$ is the dual basis to $\{w_1, \ldots, w_n\}$. If $(b_1 \cdots b_n)$ are the coordinates of $f \in V^*$ with respect to $\{g_1, \ldots, g_n\}$, then

$$f = (b_1 \cdots b_n) \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix} = (b_1 \cdots b_n) A^{-1} \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = (a_1 \cdots a_n) \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix},$$

so $(b_1 \cdots b_n)A^{-1} = (a_1 \cdots a_n)$, i.e., $(b_1 \cdots b_n) = (a_1 \cdots a_n)A$, is the transformation law for the coordinates of f with respect to the two dual bases $\{f_1, \ldots, f_n\}$ and $\{g_1, \ldots, g_n\}$.

Linear Transformations

Examples:

(1) Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation. For $1 \leq j \leq n$, let

$$t_j = \begin{pmatrix} t_{ij} \\ \vdots \\ t_{mj} \end{pmatrix} = T(e_j) \in \mathbb{F}^m.$$

If $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n$, $T(x) = T(\Sigma x_j e_j) = \Sigma x_j T(e_j)$, i.e.,

$$T(x) = (t_1 \cdots t_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & & \vdots \\ t_{m1} & \cdots & t_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

So every linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$ is given by multiplication by a matrix in $\mathbb{F}^{m\times n}$.

(2) One can construct linear transformations $G : \mathbb{F}^\infty \to \mathbb{F}^\infty$ by matrix multiplication. Let

$$T = \begin{pmatrix} t_{11} & t_{12} & \cdots \\ t_{21} & \ddots & \\ \vdots & & \end{pmatrix}$$

be an infinite matrix for which each row has only finitely many nonzero entries. In forming Tx for $x = \begin{pmatrix} x_1 \\ \vdots \end{pmatrix} \in \mathbb{F}^\infty$, each entry in Tx is given by a finite sum, so Tx makes sense and T clearly defines a linear transformation from $\mathbb{F}^\infty$ to itself. However, not all linear transformations on $\mathbb{F}^\infty$ are of this form. The shift operators $(x_1, x_2, \ldots)^T \mapsto (0, x_1, x_2, \ldots)^T$ and $(x_1, x_2, \ldots)^T \mapsto (x_2, x_3, \ldots)^T$ are examples of linear transformations on $\mathbb{F}^\infty$ that cannot be written in this form.

(3) If $\sup\limits_{i,j} |t_{ij}| < \infty$ and $x \in \ell^1$, then for each i, $\sum_{j=1}^\infty |t_{ij}x_j| \le \sup\limits_{i,j} |t_{ij}| \sum_{j=1}^\infty |x_{ij}|$. It follows that matrix multiplication Tx defines a linear transformation $T : \ell^1 \to \ell^\infty$.

(4) There are many ways linear transformations arise on function spaces, e.g.,

(a) Let $k \in C([c, d] \times [a, b])$ where $[a, b], [c, d]$ are closed bounded intervals. Define the linear transformation $L : C[a, b] \to C[c, d]$ by $L(u)(x) = \int_a^b k(x, y)u(y)dy$. L is called an *integral operator* and $k(x, y)$ is called its *kernel*.

(b) Let $m \in C[a, b]$. Then $L(u)(x) = m(x)u(x)$ defines a *multiplier operator* L on $C[a, b]$.

(c) Let $g : [c, d] \to [a, b]$. Then $L(u)(x) = u(g(x))$ defines a *composition operator* $L : C[a, b] \to C[c, d]$.

(d) $u \mapsto u'$ defines a *differential operator* $L : C^1[a, b] \to C[a, b]$.

Matrices and Basis Transformations

Suppose V, W are finite-dimensional with bases $\{v_1, \ldots, v_n\}$, $\{w_1, \ldots, w_m\}$, respectively and suppose $L : V \to W$ is linear. For $1 \leq j \leq n$, we can write $Lv_j = \sum_{i=1}^m t_{ij} w_i$. The matrix

$$T = \begin{pmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & & \vdots \\ t_{m1} & \cdots & t_{mn} \end{pmatrix} \in \mathbb{F}^{m \times n}$$

is called the matrix of L with respect to the basis $\mathcal{B}_1 = \{v_i, \ldots, v_n\}$, $\mathcal{B}_2 = \{w_i, \ldots, w_m\}$ (H-J writes $T = {}_{\mathcal{B}_2}[L]_{\mathcal{B}_1}$). If $v \in V$, $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ are the coordinates of v with respect to $\mathcal{B}_1$, and $\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$ are the coordinates of Lv with respect to $\mathcal{B}_2$, then

$$\sum_{i=1}^m y_i w_i = Lv = L\left(\sum_{j=1}^n x_j v_j\right) = \sum_{i=1}^m \left(\sum_{j=1}^n t_{ij} x_j\right) w_i,$$

so for $1 \leq i \leq m$, $y_i = \sum_{j=1}^n t_{ij} x_j$, i.e. $y = Tx$. Also, $L(v_1 \cdots v_n) = (w_1 \cdots w_n)T$. Now let $\mathcal{B}_1' = \{v_1', \ldots, v_n'\}$ and $\mathcal{B}_2' = \{w_1', \ldots, w_m'\}$ be different bases for V, W, respectively, with change-of-bases matrices $A \in \mathbb{F}^{n \times n}$, $B \in \mathbb{F}^{m \times m}$:

$$(v_1' \cdots v_n') = (v_1 \cdots v_n)A \quad \text{and} \quad (w_1' \cdots w_m') = (w_1 \cdots w_m)B.$$

Then

$$L(v_1' \cdots v_n') = (w_1 \cdots w_n)TA = (w_1' \cdots w_m')B^{-1}TA,$$

so the matrix of L in the new bases is

$${}_{\mathcal{B}_2'}[L]_{\mathcal{B}_1'} = B^{-1}TA = {}_{\mathcal{B}_2'}[I]_{\mathcal{B}_2\mathcal{B}_2}[L]_{\mathcal{B}_1\mathcal{B}_1}[I]_{\mathcal{B}_1'}.$$

In particular, if $W = V$, $\mathcal{B}_2 = \mathcal{B}_1$, and $\mathcal{B}_2' = \mathcal{B}_1'$, then $B = A$, so the matrix of L in the new basis is $A^{-1}TA$. This matrix is said to be *similar* to T. The transformation $T \mapsto A^{-1}TA$ is said to be a *similarity transformation* of A. A similarity transformation of a matrix corresponds to the representation of the same linear transformation with respect to different bases.

Linear transformations can be studied abstractly or in terms of matrix representations. For $L : V \to W$, the range $\mathcal{R}(L)$, null space $\mathcal{N}(L)$ (or kernel $\ker(L)$), $\operatorname{rank}(L) = \dim(\mathcal{R}(L))$, etc., can be defined directly in terms of L, or in terms of matrix representations. If $T \in \mathbb{F}^{n \times n}$ is the matrix of $L : V \to V$ in some basis, it is easiest to define $\det L = \det T$ and $\operatorname{tr} L = \operatorname{tr} T$. Since $\det(A^{-1}TA) = \det T$ and $\operatorname{tr}(A^{-1}TA) = \operatorname{tr}(T)$, these are independent of the choice of basis.

Vector Spaces of Linear Transformations

Let V, W be vector spaces. If $L_1 : V \to W$, $L_2 : V \to W$ are linear, define $\alpha_1 L_1 + \alpha_2 L_2 : V \to W$ for $\alpha_1, \alpha_2 \in \mathbb{F}$ by $(\alpha_1 L_1 + \alpha_2 L_2)v = \alpha_1 L_1(v) + \alpha_2 L_2(v)$; so the space of all linear transformations from V to W is naturally a vector space over $\mathbb{F}$. If V, W are finite-dimensional, we denote this vector space by $\mathcal{B}(V, W)$; in the infinite-dimensional case, we will use this notation to mean all bounded linear transformations (to be defined) from V to W with respect to norms on V, W.

Remark. For normed linear spaces, a linear operator is a bounded linear operator iff it is continuous iff it is uniformly continuous. When V, W are finite dimensional normed linear spaces, every linear transformation from V to W is continuous, and is therefore a bounded linear transformation.

If V, W have dimensions n, m, respectively, then $\mathcal{B}(V, W)$ is isomorphic to $\mathbb{F}^{m \times n}$, so it has dimension nm. When $V = W$, we denote $\mathcal{B}(V, V)$ by $\mathcal{B}(V)$. Since the composition $M \subset L : V \to U$ of linear transformations $L : V \to W$ and $M : W \to U$ is also linear, $\mathcal{B}(V)$ is naturally an algebra with composition as the multiplication operation.

Projections

Suppose W_1, W_2 are subspaces of V and $V = W_1 \oplus W_2$. Then we say W_1 and W_2 are *complementary subspaces.* Any $v \in V$ can be written uniquely as $v = w_1 + w_2$ with $w_1 \in W_1$, $w_2 \in W_2$. So we can define maps $P_1 : V \to W_1$, $P_2 : V \to W_2$ by $P_1 v = w_1$, $P_2 v = w_2$. It is easy to check that P_1, P_2 are linear. We usually regard P_1, P_2 as mapping V into itself (as $W_1 \subset V$, $W_2 \subset V$). P_1 is called the *projection onto* W_1 *along* W_2 (similarly P_2 is the projection of W_2 along W_1). It is important to note that P_1 is not determined solely by the subspace $W_1 \subset V$, but also depends on the choice of the complementary subspace W_2. Since a linear transformation is determined by its restrictions to direct summands of its domains, P_1 is uniquely characterized as that linear transformation on V which satisfies

$$P_1\Big|_{W_1} = I\Big|_{W_1} \quad \text{and} \quad P_1\Big|_{w_2} = 0.$$

It follows easily that

$$P_1^2 = P_1, \quad P_2^2 = P_2, \quad P_1 + P_2 = I, \quad \text{and} \quad P_1 P_2 = P_2 P_1 = 0.$$

In general, an element q of an algebra is call *idempotent* if $q^2 = q$. If $P : V \to V$ is a linear transformation and P is idempotent, then P is a projection in the above sense: it is the projection onto $\mathcal{R}(P)$ along $\mathcal{N}(P)$.

This discussion extends to the case in which $V = W_1 \oplus \cdots \oplus W_m$ for subspaces W_i. We can define projections $P_i : V \to W_i$ in the obvious way: P_i is the projection onto W_i along $W_1 \oplus \cdots \oplus W_{i-1} \oplus W_{i+1} \oplus \cdots \oplus W_m$. Then

$$P_i^2 = P_i \text{ for } 1 \le i \le m, \quad P_1 + \cdots + P_m = I, \quad \text{and} \quad P_i P_j = P_j P_i = 0 \text{ for } i \ne j\ .$$

If V is finite dimensional, we say that a basis $\{w_1, \ldots, w_p, u_1, \ldots, u_q\}$ for $V = W_1 \oplus W_2$ is *adapted to the decomposition* $W_1 \oplus W_2$ if $\{w_1, \ldots, w_p\}$ is a basis for W_1 and $\{u_1, \ldots, u_q\}$

is a basis for W_2. With respect to such a basis, the matrix representations of P_1 and P_2 are

$$\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}, \text{ where the block structure is } \begin{bmatrix} p\times p & p\times q \\ q\times p & q\times q \end{bmatrix},$$

$$\text{abbreviated} \quad \begin{matrix} & p & q \\ p & * & * \\ q & * & * \end{matrix}.$$

Invariant Subspaces

We say that a subspace $W \subset V$ is invariant under a linear transformation $L : V \to V$ if $L(W) \subset W$. If V is finite dimensional and $\{w_1, \ldots, w_p\}$ is a basis for W which we complete to some basis $\{w_1, \ldots, w_p, u_1, \ldots, u_q\}$ of V, then W is invariant under L iff the matrix of L in this basis is of the form

$$\begin{matrix} & p & q \\ p & * & * \\ q & 0 & * \end{matrix},$$

i.e., block upper-triangular.

We say that $L : V \to V$ preserves the decomposition $W_1 \oplus \cdots \oplus W_m = V$ if each W_i is invariant under L. In this case, L defines linear transformations $L_i : W_i \to W_i$, $1 \le i \le m$, and we write $L = L_1 \oplus \cdots \oplus L_m$. Clearly L preserves the decomposition iff the matrix T of L with respect to an adapted basis is of block diagonal form

$$T = \begin{bmatrix} T_1 & & & 0 \\ & T_2 & & \\ & & \ddots & \\ 0 & & & T_m \end{bmatrix},$$

where the T_i's are the matrices of the L_i's in the bases of the W_i's.

Nilpotents

A linear transformation $L : V \to V$ is called *nilpotent* if $L^r = 0$ for some $r > 0$. A basic example is a shift operator on $\mathbb{F}^n$: define $Se_i = 0$, and $Se_i = e_{i-1}$ for $2 \le i \le n$. The matrix of S is

$$S = S_n = \begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & & 1 \\ 0 & & & 0 \end{bmatrix} \in \mathbb{F}^{n\times n}.$$

Note that S^m shifts by m : $S^m e_i = 0$ for $1 \le i \le m$, and $S^m e_i = e_{i-m}$ for $m+1 \le i \le n$. Thus $S^n = 0$. For $1 \le m \le n-1$, the matrix $(S_n)^m$ of S^m is zero except for 1's on the m^{th}

super diagonal (i.e., the ij elements for $j = i + m$ $(1 \le i \le n - m)$ are 1's):

$$S^m = (S_n)^m = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 \\ & \ddots & \ddots & \ddots & 1 \\ & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 0 \end{bmatrix} \quad \begin{array}{l} \longleftarrow (1, m+1) \text{ element} \\ \longleftarrow (n-m, n) \text{ element.} \\ \\ \\ \end{array}$$

Note, however that the shift operator on $\mathbb{F}^\infty$: $Se_i = e_{i-1}$ $(i \ge 2)$, $Se_1 = 0$, is *not* nilpotent.

Structure of Nilpotent Operators in Finite Dimensions

Let V be finite dimensional and $L : V \to V$ be nilpotent. We will show that there is a basis for V in which L is a direct sum of shift operators. This decomposition results from a direct sum decomposition of both the domain and range of L. The basic idea is to build the decomposition by using the structure of the subspaces $\mathcal{N}(L^k)$ in the domain of L and the structure of the subspaces $\mathcal{R}(L^k)$ in the range of L. This is a key step in showing that every matrix is similar to a matrix in Jordan form.

Since L is nilpotent there is an integer r such that $L^r = 0 \ne L^{r-1}$. The proof proceeds by successively considering the subspaces $\mathcal{R}(L^{r-j})$ starting with $j = 1$ and decomposing along these subspaces. Let $v_1, \ldots, v_{\ell_1}$ be a basis for $\mathcal{R}(L^{r-1})$, and for $1 \le i \le \ell_1$, choose $w_i \in V$ for which

$$v_i = L^{r-1} w_i .$$

Observe that

$$V = \mathcal{N}(L^r) = \mathcal{N}(L^{r-1}) \oplus \operatorname{span}\{w_1, \ldots, w_{\ell_1}\}. \tag{1}$$

We claim that the set

$$\mathcal{S}_1 = \{L^{r-1}w_1, L^{r-2}w_1, \ldots, w_1,\ L^{r-1}w_2, L^{r-2}w_2, \ldots, w_2,\ \ldots,\ L^{r-1}w_{\ell_1}, L^{r-2}w_{\ell_1}, \ldots, w_{\ell_1}\}$$

is linearly independent (note the cyclic nature of the decomposition already appearing in the description of $\mathcal{S}_1$). Indeed, suppose

$$\sum_{i=1}^{\ell_1} \sum_{k=0}^{r-1} c_{ik} L^k w_i = 0\ .$$

Apply L^{r-1} to obtain

$$\sum_{i=1}^{\ell_1} c_{i0} L^{r-1} w_i = 0,$$

i.e.

$$\sum_{i=1}^{\ell_1} c_{i0} v_i = 0, \text{ so } c_{i0} = 0 \text{ for } 1 \le i \le \ell_1.$$

Now apply L^{r-2} to the double sum to obtain

$$0 = \sum_{i=1}^{\ell_1} c_{i1} L^{r-1} w_i = \sum_{i=1}^{\ell_1} c_{i1} v_i,$$

so $c_{i1} = 0$ for $1 \leq i \leq \ell_1$. Successively applying lower powers of L shows that all $c_{ik} = 0$.

Observe that for $1 \leq i \leq \ell_1$, $\operatorname{span}\{L^{r-1}w_i, L^{r-2}w_i, \ldots, w_i\}$ is invariant under L, and L acts by shifting these vectors. It follows that on $\operatorname{span}(\mathcal{S}_1)$, L is the direct sum of ℓ_1 copies of the $(r \times r)$ shift S_r, and in the basis

$$\{L^{r-1}w_1, L^{r-2}w_1, \ldots, w_1,\ L^{r-1}w_2, L^{r-2}w_2, \ldots, w_2,\ \ldots,\ L^{r-1}w_{\ell_1}, L^{r-2}w_{\ell_1}, \ldots, w_{\ell_1}\}$$

for $\operatorname{span}(\mathcal{S}_1)$, L has the matrix

$$\begin{bmatrix} S_r & & 0 \\ & \ddots & \\ 0 & & S_r \end{bmatrix}.$$

In general, $\operatorname{span}(\mathcal{S}_1)$ need not be all of V, so we aren't done.

We know that $\{L^{r-1}w_1, \ldots, L^{r-1}w_{\ell_1}\}$ is a basis for $\mathcal{R}(L^{r-1})$, and that

$$\text{(2)} \qquad \{L^{r-1}w_1, \ldots, L^{r-1}w_{\ell_1}, L^{r-2}w_1, \ldots, L^{r-2}w_{\ell_1}\}$$

are linearly independent vectors in $\mathcal{R}(L^{r-2})$ (indeed, we showed that all of the vectors in $\mathcal{S}_1$ were linearly independent, and

$$L^{r-1}w_i = L^{r-2}(Lw_i) \in \mathcal{R}(L^{r-2})\ (1 \leq i \leq \ell_1)).$$

Complete (2) to a basis of $\mathcal{R}(L^{r-2})$, if necessary, by adding vectors $\widetilde{u}_1, \ldots, \widetilde{u}_{\ell_2}$. As before, choose $\widetilde{w}_{\ell_1+j}$ for which

$$L^{r-2}\widetilde{w}_{\ell_1+j} = \widetilde{u}_j\ (1 \leq j \leq \ell_2).$$

We now further refine the direct sum decomposition in (1) using the vectors $\widetilde{w}_{\ell_1+j}$ and the subspace $\mathcal{N}(L^{r-2})$ to construct a direct sum decomposition of $\mathcal{N}(L^{r-1})$. However, the span of the vectors $\widetilde{w}_{\ell_1+j}$ may not be contained in the subspace $\mathcal{N}(L^{r-1})$ which would make this refined direct sum decomposition impossible to construct. We get around this problem by replacing the vectors $\widetilde{w}_{\ell_1+j}$ $(1 \leq j \leq \ell_2)$ by vectors in $\mathcal{N}(L^{r-1})$. This is done by projecting each $\widetilde{w}_{\ell_1+j}$ onto $\mathcal{N}(L^{r-1})$ along $\mathcal{R}(L^{r-1})$ in the following way. Recall that the vectors $v_i = Lw_i$ $(1 \leq i \leq \ell_1)$ were chosen to form a basis for $\mathcal{R}(L^{r-1})$. Now since

$$L\widetilde{u}_j = L^{r-1}\widetilde{w}_{\ell_1+j} \in \mathcal{R}(L^{r-1}),$$

there exist coefficients $a_{ij} \in \mathbb{F}$ $(1 \leq i \leq \ell_1)(1 \leq j \leq \ell_2)$ such that

$$L^{r-1}\widetilde{w}_{\ell_1+j} = \sum_{i=1}^{\ell_1} a_{ij} L^{r-1} w_i\ .$$

Set

$$w_{\ell_1+j} = \widetilde{w}_{\ell_1+j} - \sum_{i=1}^{\ell_1} a_{ij} w_i \text{ and } u_j = L^{r-2} w_{\ell_1+j}\quad (1 \leq j \leq \ell_2).$$

Replacing the $\widetilde{u}_j$'s by the u_j's still gives a basis of $\mathcal{R}(L^{r-2})$ as above (exercise). Clearly

$$L^{r-1}w_{\ell_1+j} = 0 \text{ for } 1 \leq j \leq \ell_2.$$

Observe that now have the direct sum decomposition

$$\mathcal{N}(L^{r-1}) = \mathcal{N}(L^{r-2}) \oplus \operatorname{span}\{Lw_1, \ldots, Lw_{\ell_1}, w_{\ell_1+1}, \ldots, w_{\ell_1+\ell_2}\}$$

refining that given in (1). We also have a basis for $\mathcal{R}(L^{r-2})$ of the form

$$\{L^{r-1}w_1, \ldots, L^{r-1}w_{\ell_1}, L^{r-2}w_1, \ldots, L^{r-2}w_{\ell_1}, L^{r-2}w_{\ell_1+1}, \ldots, L^{r-2}w_{\ell_1+\ell_2}\}$$

for which

$$L^{r-1}w_{\ell_1+j} = 0 \ (1 \leq j \leq \ell_2).$$

We now repeat the argument given above for the set $\mathcal{S}_1$, but for the set

$$\begin{aligned} \mathcal{S}_2 \;=\; \{ \; & L^{r-2}w_{\ell_1+1}, L^{r-3}w_{\ell_1+1}, \ldots, w_{\ell_1+1}, \\ & L^{r-2}w_{\ell_1+2}, L^{r-3}w_{\ell_1+2}, \ldots, w_{\ell_1+2}, \\ & \ldots, \\ & L^{r-2}w_{\ell_1+\ell_2}, L^{r-3}w_{\ell_1+\ell_2}, \ldots, w_{\ell_1+\ell_2}\}. \end{aligned}$$

This gives that $\mathcal{S}_1 \cup \mathcal{S}_2$ is linearly independent, and L acts on $\operatorname{span}(\mathcal{S}_2)$ as a direct sum of ℓ_2 copies of the $(r-1)\times(r-1)$ shift S_{r-1}. We can continue this argument, decreasing r one at a time and end up with a basis of $\mathcal{R}(L^0) = V$ in which L acts as a direct sum of shift operators:

$$L = \overbrace{S_r \oplus \cdots \oplus S_r}^{\ell_1} \oplus \overbrace{S_{r-1} \oplus \cdots \oplus S_{r-1}}^{\ell_2} \oplus \cdots \oplus \overbrace{S_1 \oplus \cdots \oplus S_1}^{\ell_r} \qquad (\text{Note: } S_1 = 0 \in \mathbb{F}^{1\times 1})$$

Remarks.

(1) For $1 \leq j$, let $k_j = \dim(\mathcal{N}(L^j))$. It follows easily from the above that $0 \leq k_1 < k_2 < \cdots < k_r = k_{r+1} = k_{r+2} = \cdots = n$, and thus $r \leq n$.

(2) The structure of L is determined by knowing r and $\ell_1, \ldots, \ell_r$. These, in turn, are determined by knowing $k_1, \ldots, k_n$.

Exercise: express $\ell_1, \ldots, \ell_r$ in terms of $k_1, \ldots, k_n$.

(3) General facts about nilpotent transformations follow from this normal form. For example, if $\dim V = n$ and $L : V \to V$ is nilpotent, then

- (i) $L^n = 0$
- (ii) $\operatorname{tr} L = 0$
- (iii) $\det L = 0$
- (iv) $\det(I + L) = 1$
- (v) for any $\lambda \in \mathbb{F}$, $\det(\lambda I - L) = \lambda^n$

Dual Transformations

Recall that if V and W are finite dimensional vector spaces, we denote by V^* and $\mathcal{B}(V,W)$ the dual space of V and the space of linear transformations from V to W, respectively. In the infinite dimensional case we will reserve this notation for the bounded linear functionals and transformations with respect to norms on V and W. So we now introduce the notation V' for the algebraic dual of V (*all* linear functionals on V), and $\mathcal{L}(V,W)$ for the space of *all* linear transformations from V to W.

Let $L \in \mathcal{L}(V,W)$. We define the *dual*, or *adjoint* transformation $L^* : W' \to V'$ by $(L^*g)(v) = g(Lv)$ for $g \in W'$, $v \in V$. Clearly $L \mapsto L^*$ is a linear transformation from $\mathcal{L}(V,W)$ to $\mathcal{L}(W',V')$, and $(L \circ M)^* = M^* \circ L^*$ if $M \in \mathcal{L}(U,V)$.

When V, W are finite dimension and we choose bases for V and W along with corresponding dual bases. Using these bases we can represent vectors in V, W, V^*, W^* by their coordinate vectors

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},$$

$$a = (a_1 \cdots a_n), \quad \text{and} \quad b = (b_1 \cdots b_m),$$

respectively. The linear operator $L \in \mathcal{L}(V,W)$ can then be represented by a matrix $T \in \mathbb{F}^{m \times n}$ for which $y = Tx$. Hence given $g \in W'$ having coordinates $b = (b_1 \cdots b_m)$ with respect to the dual basis, we get

$$g(Lv) = (b_1 \cdots b_m) T \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

so L^*g has coordinates $(a_1 \cdots a_n) = (b_1 \cdots b_m)T$. Thus L is represented by left-multiplication by T on column vectors, and L^* is represented by right-multiplication by T on row vectors. Another common convention is to represent the dual coordinate vectors also as columns; taking the transpose in the above gives

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = T^T \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix},$$

that is L^* can also be represented through left-multiplication by T^T on column vectors. (T^T is the transpose of T: $(T^T)_{ij} = t_{ji}$.)

We can take the dual of V' to obtain V''. There is a natural inclusion $V \to V''$: if $v \in V$, then $f \mapsto f(v)$ defines a linear functional on V'. This map is injective since if $v \neq 0$, there is an $f \in V'$ for which $f(v) \neq 0$. (Proof: Complete $\{v\}$ to a basis for V and take f to be the first vector in the dual basis.)

We identify V with its image, so we can regard $V \subset V''$. If V is finite dimensional, then $V = V''$ since $\dim V = \dim V' = \dim V''$. If V is infinite dimensional, however, then there are elements of V'' which are not in V.

If $S \subset V$ is a subset, we define the annihilator $S^\perp \subset V'$ by

$$S^\perp = \{f \in V' : (\forall\, v \in S)\ f(v) = 0\}.$$

Clearly $S^\perp = (\text{span}(S))^\perp$. Now $S^{\perp\perp} \subset V''$, and if $\dim V < \infty$, we can identify $V'' = V$ as above.

Proposition. If $\dim V < \infty$, then $S^{\perp\perp} = \text{span}(S)$.

Proof. It follows immediately from the definition that $\text{span}(S) \subset S^{\perp\perp}$. To show $S^{\perp\perp} \subset \text{span}(S)$, assume WLOG that S is a subspace. We claim that if W is an m-dimensional subspace of V and $\dim V = n$, then $\dim W^\perp = \text{codim} W = n - m$: choose a basis $\{w_1, \ldots, w_m\}$ for W, complete it to a basis $\{w_1, \ldots, w_{m+1}, \ldots, w_n\}$ for V; then clearly the dual basis vectors $\{f_{m+1}, \ldots, f_n\}$ are a basis for $W^\perp$, so $\dim W^\perp = n - m$. Hence $\dim S^{\perp\perp} = n - \dim S^\perp = n - (n - \dim S) = \dim S$, and we know $S \subset S^{\perp\perp}$. □

In complete generality, we have

Proposition. Suppose $L \in \mathcal{L}(V, W)$. Then $\mathcal{N}(L^*) = \mathcal{R}(L)^\perp$.

Proof. Clearly both are subspaces of W'. Let $g \in W'$. Then $g \in \mathcal{N}(L^*) \iff L^* g = 0 \iff (\forall v \in V)\ (L^*g)(v) = 0 \iff (\forall\, v \in V)\ g(Lv) = 0 \iff g \in \mathcal{R}(L)^\perp$. □

We are often interested in identifying $\mathcal{R}(L)$ for some $L \in \mathcal{L}(V, W)$. In the finite-dimensional case, this amounts to determining those $w \in W$ for which $\exists v \in V$ satisfying $Lv = w$; choosing bases of V, W and coordinate vectors $x \in \mathbb{F}^n$, $y \in \mathbb{F}^m$ for v, w and letting T be the matrix of L, this amounts to determining those $y \in \mathbb{F}^m$ for which the linear system $Tx = y$ can be solved. Combining the two Propositions above, we see that if $\dim W < \infty$, then $\mathcal{R}(L) = \mathcal{N}(L^*)^\perp$. Thus $\exists\, v \in V$ satisfying $Lv = w$ iff $g(w) = 0$ for all $g \in W^*$ for which $L^*g = 0$. In terms of matrices, $Tx = y$ is solvable iff

$$(b_1 \cdots b_m) \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = 0$$

for all $(b_1 \cdots b_m)$ for which $(b_1 \cdots b_m)T = 0$, or equivalently, $T^T \begin{pmatrix} b_1 \\ \cdots \\ b_m \end{pmatrix} = 0$. These are often called the *compatibility conditions* for solving the linear system $Tx = y$.

Bilinear Forms

A function $\varphi : V \times V \to \mathbb{F}$ is called a *bilinear form* if it is linear in each variable separately.

Examples:

(1) For any matrix $A \in \mathbb{F}^{n\times n}$, the function $\varphi(x, y) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i y_j$ is a bilinear form. In fact, all bilinear forms on $\mathbb{F}^n$ are of this form, as $\varphi(\sum x_i e_i \sum y_j e_j) =$

$\sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \varphi(e_i, e_j)$; just set $a_{ij} = \varphi(e_i, e_j)$. Similarly, for any finite-dimensional V, we can choose a basis $\{v_1, \ldots, v_n\}$; if φ is a bilinear form on V and $v = \sum x_i v_i$, $w = \sum y_j v_j$, then $\varphi(v, w) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \varphi(v_i, v_j) = x^T A y$ where $A \in \mathbb{F}^{n\times n}$ satisfies $a_{ij} = \varphi(v_i, v_j)$. A is called the *matrix of* φ with respect to the basis $\{v_1, \ldots, v_n\}$.

(2) One can also use infinite matrices $(a_{ij})_{i,j\geq 1}$ for $V = \mathbb{F}^\infty$ as long as convergence conditions are imposed. For example, if all $|a_{ij}| \leq M$, then $\varphi(x, y) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij} x_i y_j$ defines a bilinear form on ℓ^1 since $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |a_{ij} x_i y_j| \leq M(\sum_{i=1}^{\infty} |x_i|)(\sum_{j=1}^{\infty} |y_j|)$. Similarly if $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} |a_{ij}| < \infty$, then we get a bilinear form on ℓ^∞.

(3) If $f, g \in V'$, then $\varphi(x, y) = f(x)g(y)$ is a bilinear form.

(4) If $V = C[a, b]$, then

(i) for $k \in C([a, b] \times [a, b])$, $\int_a^b \int_a^b k(x, y)u(x)v(y)dxdy$

(ii) for $h \in C([a, b])$, $\int_a^b h(x)u(x)v(x)dx$

(iii) for $x_0 \in [a, b]$, $u(x_0) \int_a^b v(x)dx$

are all examples of bilinear forms.

We say that a bilinear form is *symmetric* if $(\forall\, v, w \in V)\ \varphi(v, w) = \varphi(w, v)$. In the finite-dimensional case, this corresponds to the condition that the matrix A be symmetric, i.e., $A = A^T$, or $(\forall\, i, j)\ a_{ij} = a_{ji}$.

Sesquilinear Forms

When $\mathbb{F} = \mathbb{C}$, we will more often use sesquilinear forms: $\varphi : V \times V \to \mathbb{C}$ is called *sesquilinear* if φ is linear in the first variable and conjugate-linear in the second variable, i.e.,

$$\varphi(v, \alpha_1 w_1 + \alpha_2 w_2) = \bar{\alpha}_1 \varphi(v_1 w_1) + \bar{\alpha}_2 \varphi(v, w_2).$$

For example, on $\mathbb{C}^n$ all sesquilinear forms are of the form $\varphi(z, w) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} z_i \bar{w}_j$ for some $A \in \mathbb{C}^{n\times n}$. To be able to discuss bilinear forms over $\mathbb{R}$ and sesquilinear forms over $\mathbb{C}$ at the same time, we will speak of a sesquilinear form over $\mathbb{R}$ and mean just a bilinear form over $\mathbb{R}$. A sesquilinear form is said to be Hermitian-symmetric (or sometimes just Hermitian) if $(\forall\, v, w \in V)\ \varphi(v, w) = \overline{\varphi(w, v)}$ (when $\mathbb{F} = \mathbb{R}$, we say the form is symmetric). This corresponds to the condition that $A = A^H$ where $A^H = \bar{A}^T$ (i.e., $(A^H)_{ij} = \overline{A_{ji}}$ is the Hermitian transpose (or conjugate transpose) of A when $\mathbb{F} = \mathbb{C}$ (in which case the matrix $A \in \mathbb{C}^{n\times n}$ is called *Hermitian*), or the condition $A = A^T$ (i.e., A is symmetric) when $\mathbb{F} = \mathbb{R}$.

To a sesquilinear form, we can associate the quadratic form $\varphi(v, v)$. We say that φ is nonnegative (or positive semi-definite) if $(\forall\, v \in V)\ \varphi(v, v) \geq 0$, and that φ is positive (or positive definite) if $\varphi(v, v) > 0$ for all $v \neq 0$ in V. By an *inner product* on V, we will mean a positive-definite Hermitian-symmetric sesquilinear form.

Examples:

(1) $\mathbb{F}^n$ with the Euclidean inner product $\langle x, y\rangle = \sum_{i=1}^{n} x_i \overline{y_i}$.

(2) Let $V = \mathbb{F}^n$, and let $A \in \mathbb{F}^{n\times n}$ be Hermitian-symmetric, and define

$$\langle x, y\rangle_A = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i\overline{y_i}.$$

The requirement that $\langle x, x\rangle_A > 0$ for $x \neq 0$ so that $\langle\cdot,\cdot\rangle_A$ is an inner product serves to define positive-definite matrices.

(3) If V is any finite-dimensional vector space, we can choose a basis and thus identify $V \cong \mathbb{F}^n$, and then transfer the Euclidean inner product to V in the coordinate of this basis. The resulting inner product depends on the choice of basis — in general there is no canonical inner product on a general vector space. With respect to the coordinates induced by a basis, any inner product on a finite-dimensional vector space V is of the form described in example (2) above.

(4) One can define an inner product on ℓ^2 by $\langle x, y\rangle = \sum_{i=1}^{\infty} x_i\overline{y_i}$. To see (from first principles) that this sum converges absolutely, apply the finite-dimensional Cauchy-Schwarz inequality to obtain

$$\sum_{i=1}^{n}|x_i\overline{y_i}| \leq \left(\sum_{i=1}^{n}|x_i|^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n}|y_i|^2\right)^{\frac{1}{2}} \leq \left(\sum_{i=1}^{\infty}|x_i|^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{\infty}|y_i|^2\right).$$

Now let $n \to \infty$ to deduce that the series $\sum_{i=1}^{\infty} x_i\overline{y_i}$ converges absolutely.

(5) The L^2-inner product on $C([a, b])$ is given by $\langle u, v\rangle = \int_a^b u(x)\overline{v(x)}dx$.

Exercise. Show that the inner product defined in Example (5) above is indeed positive definite on $C([a, b])$.

An inner product on V determines an injection $V \to V'$: if $w \in V$, define $w^* \in V'$ by $w^*(v) = \langle v, w\rangle$; since $w^*(w) = \langle w, w\rangle$ it follows that $w^* = 0 \Rightarrow w = 0$, so the map $w \mapsto w^*$ is injective. The map $w \mapsto w^*$ is *conjugate-linear* (rather than linear, unless $\mathbb{F} = \mathbb{R}$) since $(\alpha w)^* = \bar{\alpha}w^*$. The image of this map is a subspace of V'. If $\dim V < \infty$, then this map is surjective too since $\dim V = \dim V'$. In general, it is not surjective.

Let $\dim V < \infty$, and represent vectors in V as elements of $\mathbb{F}^n$ by choosing a basis. If v, w have coordinates $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, respectively, and the inner product has matrix $A \in \mathbb{F}^{n\times n}$ in this basis, then

$$w^*(v) = \langle v, w\rangle = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}\overline{y_j}\right)x_i.$$

It follows that w^* has components $b_i = \sum_{j-1}^{n} a_{ij}\overline{y_j}$ with respect to the dual basis. In terms of matrices, the map $w \mapsto w^*$ is represented by

$$\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = A\overline{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}.$$

An inner product on V allows a reinterpretation of annihilators. If $W \subset V$ is a subspace, define the orthogonal complement (read W "perp")

$$W^\perp = \{v \in V : \langle v, w \rangle = 0 (\forall\, w \in W)\}.$$

Clearly $W^\perp$ is a subspace of V. The use of the same notation that we used for the annihilator of W (a subspace of V') is justified by the observation that the image of this subspace $W^\perp$ of V under the map $V \to V'$ discussed above is precisely the annihilator of W. If $\dim V < \infty$, a dimension count and the obvious $W \cap W^\perp = \{0\}$ show that $V = W \oplus W^\perp$. So in a finite dimensional inner product space, a subspace W determines a natural complement, namely $W^\perp$. The induced projection onto W (along $W^\perp$) is called the *orthogonal projection* onto W.