Notes on Linear Algebra

Paulo Eduardo Rauber

2014

1 Vector spaces

The set of complex numbers is defined as $\mathbb{C} = \{a + bi | a, b \in \mathbb{R}\}$. By definition, $i^2 = -1$.

Addition and multiplication are defined as follows:

$$\begin{aligned} (a+bi) + (c+di) &= (a+c) + (b+d)i, \\ (a+bi)(c+di) &= (ac-bd) + (ad+bc)i. \end{aligned}$$

Addition and multiplication of complex numbers is commutative and associative. Multiplication is distributive over addition. There exists an additive identity (0) and an multiplicative identity (1).

For every $z \in \mathbb{C}$ there is an unique w such that $w + z = 0$, which we denote by $-z$. There exists, for every $z \in C$, with $z \neq 0$, a w such that $zw = 1$, which we denote by $1/z$. Division is defined as as follows: $z/w = z(1/w)$. Powers are defined as usual and have the properties $(z^m)^n = z^{mn}$ and $(zw)^m = z^m w^m$.

By convention, $\mathbb{F}$ will denote either $\mathbb{R}$ or $\mathbb{C}$ depending on context. We call $a \in F$ a scalar.

Let $z = a + bi$ be a complex number. Then $\operatorname{Re} z = a$ and $\operatorname{Im} z = b$. The complex conjugate of z is defined as $\overline{z} = a - bi$. The absolute value is defined as $|z| = \sqrt{a^2 + b^2}$. The following properties hold:

$$\begin{aligned} \operatorname{Re}(w+z) &= \operatorname{Re} w + \operatorname{Re} z, \\ \operatorname{Im}(w+z) &= \operatorname{Im} w + \operatorname{Im} z, \\ z + \overline{z} &= 2 \operatorname{Re} z, \\ z - \overline{z} &= 2(\operatorname{Im} z)i, \\ z\overline{z} &= |z|^2, \\ \overline{w+z} &= \overline{w} + \overline{z}, \\ \overline{wz} &= \overline{wz}, \\ \overline{\overline{z}} &= z, \\ |wz| &= |w||z|. \end{aligned}$$

A vector space is a set V along with operations of addition and scalar multiplication. Addition is a function that assigns an element $u + v \in V$ to each pair $u, v \in V$. Scalar multiplication is a function that assigns an element $av \in V$ for every $a \in \mathbb{F}$ and $v \in V$. A member of a vector space is called a vector. We leave the context determine whether 0 represents a vector or a scalar. The following properties must hold for V to be considered a vector space. V must be closed under addition and scalar multiplication. Addition and scalar multiplication must be commutative and associative. There must be one additive and one multiplicative identity. Scalar multiplication must be distributive over addition.

There are several examples of vector spaces: $\mathbb{R}^n$ (the set of real number n-uples), $P(\mathbb{F})$ (the set of polynomial functions with coefficients in $\mathbb{F}$) etc. Particularly interesting is $\mathbb{R}^2$, where addition and scalar multiplication can be interpreted geometrically (arrow addition and scaling).

A subspace U of V is a vector space contained in V that contains 0 and is closed over addition and scalar multiplication.

We define the sum of vector spaces as follows: $U = U_1 + U_2 + \cdots + U_n = \{u_1 + u_2 + \cdots + u_n \; | u_j \in U_j\}$. It is interesting to note that U is the smallest set that contains every U_j. We say V is a direct sum of $U_1, \cdots, U_n$ if and only if every element of v can be written uniquely as the sum of $u_1 + u_2 + \cdots + u_n$, $u_j \in U_j$, in which case we say $V = U_1 \oplus \cdots \oplus U_n$. V is a direct sum of $U_1, \cdots, U_j$ if and only if there exists only one way of writing 0 as the sum of $u_1 + \cdots + u_n$, $u_j \in U_j$. Also interesting, if $V = U \oplus W$, then $U \cap W = \{0\}$.

2 Finite-Dimensional Vector Spaces

A linear combination of a list of vectors $(v_1, \cdots, v_n)$ is a vector of the form:

$$a_1v_1 + \cdots + a_nv_n,$$

where $a_j \in F$.

The span of $(v_1, \cdots, v_n)$ is the set of all linear combinations using these vectors. The span of any list of vectors in V is a subspace of V. If a list of vectors (finite by definition) spans V, then V is said to be finite dimensional.

A list of vectors $(v_1, \cdots, v_n)$ is considered linearly independent if $a_1v_1 + \cdots + a_nv_n = 0$ implies $a_j = 0$. A list of size two is linearly independent iff one vector is not a scalar multiple of the other. If a vector is removed from a linear independent list the resulting list will be independent. The empty list has span $\{0\}$.

In a finite dimensional vector space, the length of every list of linearly independent vectors is less than or equal the length of every spanning list. A list $(v_1, \cdots v_n)$ is linearly dependent and $v_1 \neq 0$ iff there exists $j \in \{2, \cdots, m\}$ such that $v_j \in \text{span}(v_1, \cdots, v_{j-1})$ and removing v_j from this last list results in a list with the same span. If, for every positive integer n, there exists a linearly independent list of vectors in V, V is infinite dimensional. Every subspace of a finite dimensional space is finite dimensional.

A basis for a vector space V is a list of vectors in V that is linearly independent and spans V.

A list $(v_1, \cdots, v_n)$ of vectors in V is a basis of V iff every $v \in V$ can be written uniquely in the form:

$$v = a_1v_1 + \cdots + a_nv_n.$$

Every spanning list in a vector space can be reduced to a basis of the vector space. Every finite dimensional space has a basis.

Any linearly independent list of vectors in V can be extended to a basis of V.

If V is a finite dimensional vector space and U is a subspace of V, then there is a subspace W of V such that $V = U \oplus W$.

Any two basis of a vector space have the same length, which is called the dimension $\dim V$ of V.

If V is finite dimensional and U is a subspace of V, then $\dim U \leq \dim V$.

Every spanning list of V with length $\dim V$ is a basis of V. Every linearly independent list with $\dim V$ vectors in V is a basis of V.

If U_1 and U_2 are two subspaces of a finite dimensional space, then:

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

Suppose V is finite dimensional and $U_1, \cdots, U_m$ are subspaces of V such that:

$$V = U_1 + \cdots + U_m$$
$$\dim V = \dim U_1 + \cdots + \dim U_m,$$

then $V = U_1 \oplus \cdots \oplus U_m$.

3 Linear Maps

A linear map from V to W is a function $T : V \to W$ with the following properties:

$$T(u + v) = T(u) + T(v), \text{ for all } u, v \in V$$
$$T(av) = aT(v), \text{ for } a \in \mathbb{F}.$$

We denote by $\mathbb{L}(V, W)$ the set of linear maps from V to W. Two important linear maps are the map 0, defined as $0(v) = 0$ (on the left side, 0 denotes a linear map, on the right side, 0 denotes a vector) and the identity map I, defined as $I(v) = v$.

Suppose that $(v_1, \cdots, v_n)$ is a basis of V and $T \in \mathbb{L}(V, W)$. Then, by the linearity of T, $T(v) = a_1T(v_1) + \cdots + a_nT(v_n)$. This implies that $T(v_j) \in W$ define a linear transformation completely.

Let $S, T \in \mathbb{L}(V, W)$. We can define addition and scalar multiplication on $\mathbb{L}(V, W)$ as follows:

$$(S + T)(v) = S(v) + T(v)$$
$$(aT)(v) = aT(v).$$

Thus, it can easily be shown that $\mathbb{L}(V, W)$ is a vector space. We can also define the multiplication ST to be the composition of functions: $(ST)(v) = S(T(V))$. It can be shown that this operation is associative, distributive and has an additive identity.

We say the null space of a linear transformation $T : V \to W$ is the set $\operatorname{null} T = \{v \in V | T(v) = 0\}$. It can be shown that $\operatorname{null} T$ is a subspace of V. We say T is injective whenever $T(u) = T(v)$ implies $u = v$. T is also injective if and only if $\operatorname{null} T = \{0\}$.

We say the range of a linear transformation $T : V \to W$ is the set $\operatorname{range} T = \{T(v) | v \in V\}$. It can be shown that $\operatorname{range} T$ is a subspace of W. We say T is surjective whenever $\operatorname{range} T = W$.

If V is a finite dimensional vector space and $T \in \mathbb{L}(V, W)$, then $\operatorname{range} T$ is a finite dimensional subspace of W and $\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$.

If V and W are finite dimensional vector spaces and $\dim V > \dim W$ then no linear map from V to W is injective. Also, if $\dim V < \dim W$, no linear map from V to W is surjective.

Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be defined as:

$$T(x) = T(x_1, \cdots, x_n) = (\sum_{k=1}^{n} a_{1,k} x_k, \cdots, \sum_{k=1}^{n} a_{m,k} x_k).$$

Using the statements above, it can be shown that $T(x) = 0$ has more than one solution when $n > m$ (more variables than equations). Also if $n < m$ (more equations than variables), then there is not a x such that $T(x) = c$ for every choice of c for all choices of coefficients $a_{i,j}$. This relates linear maps to systems of linear equations.

An m-by-n matrix M is a rectangular array with m rows and n columns that can be represented as:

$$M = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}$$

Suppose $(v_1, \cdots, v_n)$ is a basis of V and $(w_1, \cdots, w_m)$ is a basis of W. We can write $T(v_k) = a_{1,k} w_1 + \cdots + a_{m,k} w_m$ for every $k = 1, \ldots, n$ and scalars $a_{j,k}$ for $j = 1, \ldots, m$. The m-by-n matrix formed by the a's is called the matrix of T with respect to the basis $(v_1, \cdots, v_n)$ and $(w_1, \cdots, w_m)$, which we denote by $M(T, (v_1, \cdots, v_n), (w_1, \cdots, w_m))$, or $M(T)$ when the context makes the basis clear. The values in column k are the scalars that need to be multiplied by $w_1, \ldots, w_m$ to obtain $T(v_k)$.

Unless stated otherwise, if T is a linear map from $\mathbb{F}^n$ to $\mathbb{F}^m$, the basis is the standard (where the k-th vector in the basis is 0 in every slot except the k-th).

Define matrix addition and scalar multiplication to be element-wise. The set $Mat(m, n, \mathbb{F})$ contains all m-by-n matrices over $\mathbb{F}$. It is also a vector space.

If $M(T)$ is a m-by-n matrix that can be indexed by a's and $M(S)$ is a n-by-p matrix that can be indexed by b's, then we define the m-by-p matrix $M(S)M(T)$ (indexed by c's) to have the following elements:

$$c_{j,k} = \sum_{r=1}^{n} a_{j,r} b_{r,k}.$$

Using this definition of matrix multiplication, $M(T)M(S) = M(TS)$.

Suppose $(v_1, \cdots, v_n)$ is a basis of V. If $v \in V$, then there exist unique scalars $b_1, \cdots, b_n$ such that $v = b_1 v_1 + \cdots + b_n v_n$. We denote the matrix of v as:

$$M(v) = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Using this definition, $M(T(v)) = M(T)M(v)$.

We say a linear map $T \in \mathbb{L}(V, W)$ is invertible if there exists a linear map S such that $ST = I$ (identity map on V) and $TS = I$ (identity map on W). If T is invertible, its inverse is unique and denoted by T^{-1}. A linear map is invertible if and only if it is injective and surjective. Two vector spaces are called isomorphic if there is an invertible linear map between them.

Two finite-dimensional vector spaces are isomorphic if and only if they have the same dimension. Thus, every finite-dimensional vector space is isomorphic to some F^n.

Suppose $(v_1, \cdots, v_n)$ is a basis of V and $(w_1, \cdots, w_m)$ is a basis of W. Then M is as an invertible map between $L(V,W)$ and $Mat(m,n,F)$. In other words, there is a matrix for every linear map from F^n to F^m.

If V and W are finite dimensional, then $L(V,W)$ is finite dimensional and $\dim L(V,W) = (\dim V)(\dim W)$.

A linear map from a vector space V to itself is called an operator, denoted by $L(V)$. If V is finite dimensional and $T \in L(V)$, the following are equivalent: T is injective, T is surjective, T is invertible.

4 Polynomials

A function $p : \mathbb{F} \to \mathbb{F}$ is called a polynomial with coefficients in $\mathbb{F}$ if there exist $a_0, \ldots, a_m \in \mathbb{F}$ such that:

$$p(z) = a_0 + a_1 z + \ldots + a_m z^m,$$

for all $z \in \mathbb{F}$. If p can be written in the form above and $a_m \neq 0$, we say p has degree m. We use $\mathbb{P}(\mathbb{F})$ to denote the vector space of all polynomials and $\mathbb{P}_m(\mathbb{F})$ to denote the vector space of polynomials with degree at most m. A number $\lambda \in \mathbb{F}$ is called a root of p if $p(\lambda) = 0$. The polynomial $p(z) = 0$ has degree $-\infty$. If $p \in \mathbb{P}(\mathbb{F})$ is a polynomial with degree $m \geq 1$, λ is a root of p if and only if there is $q \in \mathbb{P}(\mathbb{F})$ with degree $m - 1$ such that $p(z) = (z - \lambda)q(z)$ for all $z \in \mathbb{F}$.

A polynomial with degree $m \geq 0$ has at most m distinct roots in $\mathbb{F}$. Suppose $p(z) = a_0 + a_1 z + \ldots + a_m z^m = 0$ for every $z \in \mathbb{F}$, then $a_0 = \ldots = a_m = 0$. This implies that $(1, z, \ldots, z^m)$ is a basis for $\mathbb{P}(\mathbb{F})$.

If $p, q \in \mathbb{P}(\mathbb{F})$ and $p \neq 0$, then there exist unique $s, r \in \mathbb{P}(\mathbb{F})$ such that $q = sp + r$ and $\deg r < \deg p$.

The fundamental theorem of algebra states that every nonconstant polynomial with complex coefficients has a root.

If $p \in \mathbb{P}(\mathbb{C})$ is a nonconstant polynomial, then p has an unique factorization of the form:

$$p(z) = c(z - \lambda_1) \ldots (z - \lambda_m),$$

where $c, \lambda_1, \ldots, \lambda_m \in \mathbb{C}$.

If p is a polynomial with real coefficients and $\lambda \in \mathbb{C}$ is a root of p, then so is $\overline{\lambda}$.

Let $\alpha, \beta \in \mathbb{R}$. Then there is a polynomial factorization of the form $x^2 + \alpha x + \beta = (x - \lambda_1)(x - \lambda_2)$, with $\lambda_1, \lambda_2 \in \mathbb{R}$, if and only if $\alpha^2 \geq 4\beta$.

If $p \in \mathbb{P}(\mathbb{R})$ is a nonconstant polynomial, then p has an unique factorization of the form:

$$p(x) = c(x - \lambda_1) \ldots (x - \lambda_m)(x^2 + \alpha_1 x + \beta_1) \ldots (x^2 + \alpha_M + \beta_M),$$

where $c, \lambda_1, \ldots, \lambda_m \in \mathbb{R}$, $(\alpha_1, \beta_1), \ldots, (\alpha_M, \beta_M) \in \mathbb{R}^2$ with $\alpha_j^2 < 4\beta_j$ for each j, and m or M may be equal to 0.

There exists a polynomial $p \in \mathbb{P}_n(\mathbb{F})$ with m roots for $1 \leq m \leq n$.

Suppose that $z_1, \ldots, z_{m+1}$ are distinct elements of $\mathbb{F}$ and $w_1, \ldots, w_{m+1} \in \mathbb{F}$. Then there exists an unique polynomial $p \in \mathbb{P}(\mathbb{F})$ such that $p(z_j) = w_j$.

If p is a polynomial with degree m, then p and its derivative p' have no roots in common if and only if p has m distinct roots. Every polynomial with odd degree and real coefficients has a real root.

5 Eigenvalues and Eigenvectors

Suppose $V = U_1 \oplus \cdots \oplus U_m$. It is useful to consider the behavior of an operator $T \in \mathcal{L}(V)$ on a subspace U_j of V. We denote by $T|_{U_j}$ the restriction of T to the domain U_j. We say that U_j is invariant under T if $T|_{U_j}$ is an operator on U_j. In other words, if for any $u \in U_j$, $T(u) \in U_j$. It is easy to see that $\operatorname{null} T$ and $\operatorname{range} T$ are invariant under T.

A scalar $\lambda \in \mathcal{F}$ is called an eigenvalue of $T \in \mathcal{L}(V)$ if there exists a nonzero vector $u \in V$ such that $T(u) = \lambda u$. Therefore, T has an one dimensional invariant subspace if and only if T has an eigenvalue. Since the equation $T(u) = \lambda u$ is equivalent to $T(u) - \lambda I(u) = 0$, λ is an eigenvalue of T if and only if $T(u) - \lambda I(u)$ is not injective (maps a nonzero vector to 0), surjective or not invertible.

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathcal{F}$ is an eigenvalue of T. A vector $u \in V$ is called an eigenvector of T corresponding to λ if $T(u) = \lambda u$. Thus, the set of eigenvectors corresponding to λ is $\operatorname{null}(T(u) - \lambda I(u))$, a subspace of V.

As an example, consider $T \in \mathcal{L}(\mathcal{F}^2)$, where $T(w,z) = (-z, w)$. If $\mathcal{F} = \mathbb{R}$, T has no eigenvalues. If $\mathcal{F} = \mathbb{C}$, its eigenvalues are i and $-i$.

Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $v_1, \ldots, v_m$ are corresponding nonzero eigenvectors. Then $(v_1, \ldots, v_m)$ is linearly independent. As a corollary, each operator on V has at most $\dim V$ distinct eigenvalues.

If $T \in \mathcal{L}(V)$, we can denote TT by T^2. In general, T^m denotes the application of T m times. We let T^0 denote I. If T is invertible, its inverse is denoted by T^{-1} and $T^{-m} = (T^{-1})^m$. From these definitions, it follows that $T^mT^n = T^{m+n}$ and $(T^m)^n = T^{mn}$ for arbitrary integers n and m as long as T is invertible and non-negative integers if T is not invertible.

If $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathcal{F})$ is a polynomial given by $p(z) = a_0 + a_1 z + \ldots + a_m z^m$ for $z \in \mathcal{F}$, we define $p(T)$ as the operator $p(T) = a_0 I + a_1 T + \ldots + a_m T^m$. The function from $\mathcal{P}(\mathcal{F})$ to $\mathcal{L}(V)$ is linear. If p and q are polynomials, $(pq)(z) = p(z)q(z)$. It is also true that $(pq)(T) = p(T)q(T)$ for any $T \in \mathcal{L}(V)$.

Every operator on a finite-dimensional, nonzero, complex vector space has an eigenvalue.

If V is a complex vector space, there exists a basis with respect to which the matrix of T has has zeros everywhere in the first column except in the first entry.

The diagonal of a square matrix consists of the entries along the straight line from the upper left corner to the bottom right corner $(a_{1,1}, \ldots, a_{n,n})$. A matrix is called upper triangular if all the entries bellow the diagonal are zero ($*$ denotes irrelevant values, 0 represents all the entries bellow the diagonal):

$$\begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}.$$

Suppose $T \in \mathcal{L}(V)$ and $(v_1, \ldots, v_n)$ is a basis of V. Then the following are equivalent: the matrix of T wrt $(v_1, \ldots, v_n)$ is upper triangular, $T(v_k) \in \operatorname{span}(v_1, \ldots, v_k)$ and $\operatorname{span}(v_1, \ldots, v_k)$ is invariant under T.

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.

Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix wrt some basis of V. Then T is invertible if and only if all the entries on the diagonal of that upper-triangular matrix are nonzero.

Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix wrt some basis of V. Then the eigenvalues of T consist precisely of the entries in the diagonal of that upper-triangular matrix.

A diagonal matrix is a square matrix that is 0 everywhere except possibly the diagonal. An operator $T \in \mathcal{L}(V)$ has a diagonal matrix wrt some basis of V if and only if V has a basis consisting of eigenvectors of T. If $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T has a diagonal matrix wrt some basis of V.

Suppose $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T. Then the following are equivalent:

- T has a diagonal matrix wrt some basis of V;
- T has a basis consisting of eigenvectors of T;
- there exist one-dimensional subspaces $U_1, \ldots, U_n$ of V, each invariant under T, such that $V = U_1 \oplus \ldots \oplus U_n$;
- $V = \operatorname{null}(T - \lambda_1 I) \oplus \ldots \oplus \operatorname{null}(T - \lambda_m I)$;
- $\dim V = \dim \operatorname{null}(T - \lambda_1 I) + \ldots + \dim \operatorname{null}(T - \lambda_m I)$.

Every operator on a finite-dimensional, nonzero, real vector space has an invariant subspace of dimension 1 or 2.

Suppose U and W are subspaces of V with $V = U \oplus W$. Each vector $v \in V$ can be written uniquely in the form $v = u + w$. With this representation, define $P_{U,W} \in \mathcal{L}(V)$ by $P_{U,W}(v) = u$. This is called the projection onto U with null space W.

Every operator on an odd-dimensional real vector space has an eigenvalue.

If $T \in \mathcal{L}(V)$ is invertible and $\lambda \in \mathcal{F}\ \{0\}$, then λ is an eigenvalue of T if and only if $\frac{1}{\lambda}$ is an eigenvalue of T^{-1}.

For any $S, T \in \mathcal{L}(V)$, ST and TS have the same eigenvalues.

6 Inner-Product Spaces

The length of a vector x in $\mathbb{R}^2$ or $\mathbb{R}^3$ is called the norm of x and denoted by $||x||$. For $x = (x_1, x_2) \in \mathbb{R}^2$, $||x|| = \sqrt{x_1^2 + x_2^2}$. In general, the norm $||x||$ of $x \in R^n$ equals $\sqrt{x_1^2 + \ldots + x_n^n}$.

The dot product of $x, y \in \mathbb{R}^n$, denoted $x \cdot y$ is defined by $x \cdot y = x_1 y_1 + \ldots + x_n y_n$. Therefore, $||x||^2 = x \cdot x$. For $x, y \in \mathbb{C}^n$ the dot product is defined as $x \cdot y = x_1 \overline{y_1} + \ldots + x_n \overline{y_n}$, and the norm is defined as $||x|| = \sqrt{|x_1|^2 + \ldots + |x_n|^n}$. The dot product is a special case of a function called an inner product.

An inner-product on V is a function that takes each pair (u, v) of elements of V to a number $\langle u, v\rangle \in \mathbb{F}$ such that $\langle v, v\rangle \geq 0$ for every v; $\langle v, v\rangle = 0$ if and only if $v = 0$; $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$; $\langle av, w\rangle = a\langle v, w\rangle$ for all $v, w \in V$; and $\langle v, w\rangle = \overline{\langle w, v\rangle}$ for all $v, w \in V$. Clearly, when $\mathbb{F} = \mathbb{R}$, $\langle v, w\rangle = \langle w, v\rangle$. An inner-product space is a vector space V along with an inner-product on V. The Euclidean inner-product on $\mathbb{F}^n$ is the standard inner-product on $\mathbb{F}^n$ and is defined in the same way as the dot product.

We can define the inner product between polynomials in $\mathbb{P}_m(\mathbb{F})$ in the following way:

$$\langle p, q\rangle = \int_0^1 p(x)\overline{q(x)}dx. \tag{1}$$

From the basic properties of an inner product space V, the following can also be derived for any $a \in \mathbb{F}$ and $v, u, w \in V$: $\langle w, 0\rangle = \langle 0, w\rangle = 0$; $\langle u, v+w\rangle = \langle u, v\rangle + \langle u, w\rangle$; $\langle u, av\rangle = \overline{a}\langle u, v\rangle$.

The norm of $v \in V$ is given by $||v|| = \sqrt{\langle v, v\rangle}$. We have that $||v|| = 0$ iff $v = 0$. Also, $||av|| = |a|||v||$ for any $a \in \mathbb{F}$.

Two vectors u and v are said to be orthogonal if $\langle u, v\rangle = 0$. This also implies that $\langle v, u\rangle = 0$. The vector 0 is orthogonal to every vector, and the only to be orthogonal to itself.

If u and v are nonzero vectors in $\mathbb{R}^2$, $\langle u, v\rangle = ||u||\ ||v|| \cos\theta$, where θ is the angle between the two vectors.

The following properties have important geometrical interpretations in $\mathbb{R}^2$ and $\mathbb{R}^3$.

If $u, v \in V$ are orthogonal, then $||u+v||^2 = ||u||^2 + ||v||^2$. This is known as the Pythagorean theorem.

For any vector $u \in V$ and a given vector $v \in V$, we can write u as a scalar multiple of v plus a vector orthogonal to v in the following way:

$$u = \frac{\langle u, v\rangle}{||v||^2}v + (u - \frac{\langle u, v\rangle}{||v||^2}v).$$

If $u, v \in V$, then $|\langle u, v\rangle| \leq ||u||||v||$, and this is a equality if and only one of u, v is a scalar multiple of the other. This is known as the Cauchy-Schwarz inequality.

If $u, v \in V$, then $||u+v|| \leq ||u|| + ||v||$, and this is an equality if and only if one of u, v is a non-negative multiple of the other. This is known as the Triangle Inequality.

If $u, v \in V$, then $||u+v||^2 + ||u-v||^2 = 2(||u||^2 + ||v||^2)$. This is known as the parallelogram equality.

A list $(e_1, \dots, e_m)$ of vectors in V is orthonormal if $\langle e_j, e_k\rangle = 0$ when $j \neq k$ and $\langle e_j, e_k\rangle = 1$ when $j = k$.

If $(e_q, \dots, e_m)$ is an orthonormal list of vectors, then $||a_1e_1 + \dots a_me_m||^2 = |a_1|^2 + \dots + |a_m|^2$. Thus, every orthonormal list of vectors is linearly independent. An orthonormal basis of V is an orthonormal list of vectors in V that is also a basis of V.

The following is very useful to find the scalars to write any v as a linear combination of an orthonormal basis of V. Suppose $(e_1, \dots, e_n)$ is an orthonormal basis of V. Then $v = \langle v, e_1\rangle e_1 + \dots + \langle v, e_n\rangle e_n$. Also, $||v||^2 = |\langle v, e_1\rangle|^2 + \dots + |\langle v, e_n\rangle|^2$.

The Gram-Schmidt process can be used to find an orthonormal list of vectors given a linearly independent list. If $(v_1, \dots, v_m)$ is a linearly independent list of vectors in V, then there exists an orthonormal list $(e_1, \dots, e_m)$ of vectors in V such that $\text{span}(v_1, \dots, v_j) = \text{span}(e_1, \dots, e_j)$ for $j = 1, \dots, m$. We let $e_1 = v_1/||v_1||$, and e_j for $j > 1$ is given by the following:

$$e_j = \frac{v_j - \langle v_j, e_1\rangle e_1 - \dots - \langle v_j, e_{j-1}\rangle e_{j-1}}{||v_j - \langle v_j, e_1\rangle e_1 - \dots - \langle v_j, e_{j-1}\rangle e_{j-1}||}.$$

Every finite-dimensional inner-product space has an orthonormal basis. Also, every orthonormal list of vectors in V can be extended to an orthonormal basis of V.

Suppose $T \in \mathbb{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.

If V is a complex vector space and $T \in \mathbb{L}(V)$, then T has an upper-triangular matrix with respect to some orthonormal basis of V.

If U is a subset of V, the orthogonal complement of U, denoted by $U^\perp$, is the set of vectors that are orthogonal to every vector in U. Thus, $U^\perp = \{v \in V | \langle v, u\rangle = 0 \text{ for all } u \in U\}$. It can be shown that $U^\perp$ is always a subspace of V, $V^\perp = \{0\}$, and that $\{0\}^\perp = V$. Also, if $U_1 \subset U_2$, $U_1^\perp \supset U_2^\perp$.

If U is a subspace of V, then $V = U \oplus U^\perp$. Also, $(U^\perp)^\perp = U$.

A vector $v \in V$ can be written uniquely in the form $v = u + w$ for $u \in U$ and $w \in U^\perp$, in which case we say $P_U(v) = u$. The function P_U is an operator on V and also has the following properties: range $P_U = U$; null $P_U = U^\perp$, $v - P_U(v) \in U^\perp$ for every $v \in V$; ${P_U}^2 = P_U$; $||P_U|| \leq ||v||$ for every $v \in V$.

It is also important to note that if $(e_1, \dots, e_m)$ is an orthonormal basis of U, $P_U(v) = \langle v, e_1\rangle e_1 + \dots + \langle v, e_m\rangle e_m$.

The following proposition has led to many applications of inner-product spaces outside of pure mathematics.

Suppose U is a subspace of V and $v \in V$. Then $||v - P_U(v)|| \leq ||v - u||$ for every $u \in U$. Furthermore, if $u \in U$ and the inequality is an equality, then $u = P_U(v)$. Intuitively, this proposition means that for a given v, the nearest point in U to v is the projection of v onto U.

The general form of this minimization problem is as follows. Consider a vector space V and a subspace U. Given a vector $v \in V$, the objective is to find the nearest vector $u \in U$. The solution to this problem consists simply on finding an orthonormal base for U and projecting v onto U.

In this minimization scenario, it is particularly interesting to consider $C[a, b]$, the (infinite dimensional) vector space composed of continuous functions in an interval $[a, b]$. Given the inner product defined by Eq. 1 (pg. 6), it is easy to find the polynomial that best approximates a given continuous function by computing some definite integrals.

A linear functional on V is a linear map from V to the scalars $\mathbb{F}$. For instance, let $\varphi : \mathbb{F}^3 \to \mathbb{F}$ be defined as $\varphi(z_1, z_2, z_3) = 2z_1 + 3z_2 + 1z_3$, then φ is a linear functional on $\mathbb{F}^3$.

Suppose φ is a linear functional on V. Then there is an unique vector $v \in V$ such that $\varphi(u) = \langle u, v\rangle$ for every $u \in U$.

Let $T \in \mathbb{L}(V, W)$. The adjoint of T, denoted T^*, is defined as follows. Fix $w \in W$. Consider the linear functional that maps $v \in V$ to $\langle Tv, w\rangle$. Let T^*w be the unique vector in V such that this linear functional is given by taking inner products with T^*w. In other words, T^*w is the unique vector in V such that $\langle Tv, w\rangle = \langle v, T^*w\rangle$ for all $v \in V$.

It is possible to show that $T^* \in \mathbb{L}(W, V)$. Adjoints also has the following properties: $(S + T)^* = S^* + T^*$ for all $S, T \in \mathbb{L}(V, W)$; $(aT)^* = \overline{a}T^*$ for all $a \in \mathbb{F}$, $T \in \mathbb{L}(V, W)$; $(T^*)^* = T$ for all $T \in \mathbb{L}(V, W)$; $I^* = I$; $(ST)^* = T^*S^*$. Also, $\operatorname{null} T^* = (\operatorname{range} T)^\perp$ and $\operatorname{range} T^* = (\operatorname{null} T)^\perp$.

The conjugate transpose of a matrix is obtained by exchanging rows and columns and then taking the complex conjugate of each entry. Suppose $T \in \mathbb{L}(V, W)$. If $(e_1, \ldots, e_n)$ is an orthonormal basis of V and $(f_1, \ldots, f_m)$ is an orthonormal basis of V, then $M(T^*, (f_1, \ldots, f_n), (e_1, \ldots, e_m))$ is the conjugate transpose of $M(T, (e_1, \ldots, e_m), (f_1, \ldots, f_n))$

Let $T \in \mathbb{L}(V)$. A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if $\overline{\lambda}$ is an eigenvalue of T^*.

Let $T \in \mathbb{L}(V)$ and U be a subspace of V. Then U is invariant under T if and only if $U^\perp$ is invariant under T^*.

Let $T \in \mathbb{L}(V)$. Then T is injective if and only if T^* is surjective. Also, T^* is surjective if and only if T is injective. Also, $\dim \operatorname{range} T^* = \dim \operatorname{range} T$.

Let A be an m-by-n matrix of real numbers. The dimension of the span of the columns of A (in $\mathbb{R}^m$) equals the dimension of the span of the rows of A (in $\mathbb{R}^n$).

License

References

[1] Sheldon Axler, *Linear Algebra Done Right*. Springer, 1997.