5 Linear Subspaces of R^n

5.1 Vector spaces

Intuitive geometry ends in R^3 but its formal vector notions and concepts survive and are naturally perpetuated into higher dimensions. The language of vector spaces is at once geometrically allusive and idiomatically adept in describing general linear actions and relationships and has become a preferred mode of speech in diverse fields of linear mathematics. We shall first propose new, geometrically reminiscent, terminology for established linear algebraic facts and then interplay them into more elaborate statements of the language. A good part of the subject of vector spaces consists of fashioning a syntax.

Definition. *Let $S : v_1, v_2, \ldots, v_n$ be a set (order not important) of n vectors for which the linear combination*

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \tag{5.1}$$

is defined. The totality of vectors thus formed with arbitrary (real) numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ constitutes a vector space, say V. The n vectors of set S span V. Vector v is an element of V, $v \in V$, if and only if scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$ exist so that

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n. \tag{5.2}$$

Every vector $v \in V$ consists of a *homogeneous* linear combination of the n vectors $v_1, v_2, \ldots, v_n$ that span V, and every vector space must count the zero vector $v = o$ among its elements. Also, if $v \in V$, then necessarily $-v \in V$.

In other words, an infinite collection of vectors constitutes a linear vector space if every vector of the collection may be written as a linear homogeneous combination of a group of some n vectors— the span or generator of the space, and, conversely, if every vector generated by the linear combination belongs to the collection.

Given a set S we easily generate vectors in V with different α's, but it can be costly to decide whether or not a given vector v is in V. For the latter, a system of linear equations

$$[v_1 \ v_2 \dots \ v_n][\alpha_1 \ \alpha_2 \dots \alpha_n]^T = v \tag{5.3}$$

must be checked for a solution.

Examples and counterexamples.

1. Vectors $e_1, e_2, \dots, e_n$ span R^n. Every vector $x = [x_1 \ x_2 \ \dots \ x_n]^T$ with n (real) components is written as $x = x_1e_1 + x_2e_2 + \cdots + x_ne_n$.

2. The totality of vectors $x = [x_1 \ x_2 \ x_3]^T$, $x_1 = 1$, does not constitute a vector space since

$$x = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{5.4}$$

is not the required homogeneous linear combination; it does not contain the zero vector.

3. Vectors $x = [x_1 \ x_2 \ x_3]^T$, $x_1 + x_2 + x_3 = 0$, do constitute a vector space since after substitution

$$x = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{5.5}$$

where x_2 and x_3 are arbitrary.

4. The totality of vectors in space V excluding an element of the space does not constitute a vector space. The reader should convince himself of this fact.

5. The reader should also have no difficulty convincing himself that $x = \alpha_1^2 e_1 + \alpha_2 e_2 + \alpha_3 e_3$ is not a vector space.

Definition. *The m rows of matrix $A = A(m \times n)$, considered vectors in R^n, span a vector space called the row space of A. Similarly, the n columns of A, considered vectors in R^m, span a vector space called the column space of A.*

Theorem 5.1. *If $v \in V$ and $v' \in V$, then so is $\alpha v + \alpha' v'$.*

Proof. Since

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \quad \text{and} \quad v' = \alpha'_1 v_1 + \alpha'_2 v_2 + \cdots + \alpha'_n v_n \tag{5.6}$$

then

$$\alpha v + \alpha' v' = (\alpha\alpha_1 + \alpha'\alpha'_1)v_1 + (\alpha\alpha_2 + \alpha'\alpha'_2)v_2 + \cdots + (\alpha\alpha_n + \alpha'\alpha'_n)v_n. \tag{5.7}$$

End of proof.

Definition. *Let V and W be two vector spaces. If every $w \in W$ is also in V, then W is a subspace of V, $W \subset V$. If W is a subspace of V, and V a subspace of W, then spaces V and W are equal*, $V = W$.

To be equal, the number of vectors in the span of V need not be equal to the number of vectors in the span of W.

Theorem 5.2. *A necessary and sufficient condition for W to be a subspace of V is that the spanning set of W be in V. Spaces V and W are equal if and only if span V is in W and span W is in V.*

Proof. The condition is sufficient, for if the span of W is in V, then by Theorem 5.1 any linear combination of the spanning vectors is in V. The condition is necessary for if part of span W is not in V, then there are vectors in W not in V. Reversing the roles of V and W, we conclude that $V = W$ if and only if span V is in W and span W is in V. End of proof.

Examples.

1. Vector spaces $V : v_1, v_2, v_3$, with $v_1 = [1\ \ -2\ 1]^T, v_2 = [1\ 1\ 1]^T$, $v_3 = [1\ 0\ \ -1]^T$; and R^3 are equal vector spaces.

Obviously span V is in R^3, and we also have that

$$\begin{bmatrix}1\\0\\0\end{bmatrix} = \frac{1}{6}\begin{bmatrix}1\\-2\\1\end{bmatrix} + \frac{1}{3}\begin{bmatrix}1\\1\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix}, \quad \begin{bmatrix}0\\1\\0\end{bmatrix} = -\frac{1}{3}\begin{bmatrix}1\\-2\\1\end{bmatrix} + \frac{1}{3}\begin{bmatrix}1\\1\\1\end{bmatrix}$$

$$\begin{bmatrix}0\\0\\1\end{bmatrix} = \frac{1}{6}\begin{bmatrix}1\\-2\\1\end{bmatrix} + \frac{1}{3}\begin{bmatrix}1\\1\\1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\0\\-1\end{bmatrix}. \tag{5.8}$$

2. Vector spaces

$$V:\ \begin{bmatrix}1\\1\\1\\1\end{bmatrix},\ \begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}\quad \text{and } W:\ \begin{bmatrix}1\\0\\-1\\0\end{bmatrix},\ \begin{bmatrix}0\\1\\0\\-1\end{bmatrix} \tag{5.9}$$

are disjoint subspaces of R^4. The two vectors that span V are not in W and the two vectors that span W are not in V. No $v \in V$ is in W, and no $w \in W$ is in V.

3. Vector spaces

$$V:\ \begin{bmatrix}1\\1\\1\\1\end{bmatrix},\ \begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}\quad \text{and } W:\ \begin{bmatrix}1\\5\\1\\5\end{bmatrix},\ \begin{bmatrix}2\\-4\\2\\-4\end{bmatrix},\ \begin{bmatrix}1\\0\\1\\0\end{bmatrix} \tag{5.10}$$

are equal subspaces of R^4.

Let $V\colon v_1, v_2$ and $W\colon w_1, w_2, w_3$. According to Theorem 5.2 the two vector spaces are equal if scalars T_{ij} and T'_{ij} exist so that

$$\begin{aligned} v_1 &= T_{11}w_1 + T_{21}w_2 + T_{31}w_3 \\ v_2 &= T_{12}w_1 + T_{22}w_2 + T_{32}w_3 \end{aligned} \quad \text{and} \quad \begin{aligned} w_1 &= T'_{11}v_1 + T'_{21}v_2 \\ w_2 &= T'_{12}v_1 + T'_{22}v_2, \\ w_3 &= T'_{13}v_1 + T'_{23}v_3 \end{aligned} \tag{5.11}$$

or in compact matrix notation

$$[v_1\ v_2] = [w_1\ w_2\ w_3]T \quad \text{and} \quad [w_1\ w_2\ w_3] = [v_1\ v_2]T'. \tag{5.12}$$

We write out both equations as

$$\begin{bmatrix}1 & 2 & 1\\5 & -4 & 0\\1 & 2 & 1\\5 & -4 & 0\end{bmatrix}T = \begin{bmatrix}1 & 1\\1 & -1\\1 & 1\\1 & -1\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}1 & 1\\1 & -1\\1 & 1\\1 & -1\end{bmatrix}T' = \begin{bmatrix}1 & 2 & 1\\5 & -4 & 0\\1 & 2 & 1\\5 & -4 & 0\end{bmatrix}, \tag{5.13}$$

then routinely solve them for

$$T = \begin{bmatrix}1 & -1\\1 & -1\\-2 & 4\end{bmatrix} \quad \text{and} \quad T' = \frac{1}{2}\begin{bmatrix}6 & -2 & 1\\-4 & 6 & 1\end{bmatrix} \tag{5.14}$$

in which T' is unique but not T.

exercises

5.1.1. Is the set of all vectors $x = |\alpha|[1\ 1]^T$ for arbitrary α a vector space?

5.1.2. Is $\alpha_1[1\ 0]^T + \alpha_2^3[1\ 1]^T$ a vector space? What about $\alpha_1[1\ 0]^T + (1+\alpha_2)[1\ 1]^T$?

5.1.3. Is

$$V: \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

R^2?

5.1.4. Is $v = [1\ 0\ 1]^T$ an element of $V:\ [1\ 0\ -1]^T,\ [-1\ 2\ 1]^T$?

5.1.5. Show that the totality of vectors

$$v = \alpha_1 \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + \alpha_2 \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix} + \begin{bmatrix} -1 \\ 4 \\ -5 \end{bmatrix}$$

obtained with arbitrary α_1 and α_2 constitutes a vector space. Write a span for this space. Show on the other hand that the totality of vectors

$$v = \alpha_1 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

is not a linear vector space. Hint: look for $v = o$.

5.1.6. Show that

$$V: v_1, v_2: \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \quad \text{and} \quad W: w_1, w_2, w_3: \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

are equal subspaces of R^3.

5.2 Linear independence

Definition. *Vectors $v_1, v_2, \ldots, v_n$ of space V are linearly dependent if there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$, not all zero, so that*

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = o. \tag{5.15}$$

Vectors that are not linearly dependent are linearly independent.

Linear dependence is the extension to higher dimensions of the geometric notions of colinearity and coplanarity.

Theorem 5.3. *If vectors $v_1, v_2, \ldots, v_n$ of V are linearly dependent, then at least one of them can be written as a linear combination of the others. If vector $v \in V$ cannot be written as a linear combination of the n linearly independent vectors $v_1, v_2, \ldots, v_n$ of V, then $v, v_1, v_2, \ldots, v_n$ are linearly independent.*

Proof. By the assumption of linear dependence at least one coefficient, say α_k, is nonzero, and

$$v_k = -\frac{\alpha_1}{\alpha_k}v_1 - \frac{\alpha_2}{\alpha_k}v_2 - \cdots - \frac{\alpha_n}{\alpha_k}v_n. \tag{5.16}$$

If vector v cannot be expressed as a linear combination of the n vectors, then $\alpha v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n$ holds only for $\alpha = \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$ and the vectors are linearly independent. End of proof.

In example 3 of the previous section the span of V is linearly independent, whereas that of W is linearly dependent:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 5 \\ 1 \\ 5 \end{bmatrix} + 5\begin{bmatrix} 2 \\ -4 \\ 2 \\ -4 \end{bmatrix} - 14\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}. \tag{5.17}$$

Theorem 5.4. *Any set of vectors:*

1. Which contains the zero vector is linearly dependent.

2. Which contains a linearly dependent subset is linearly dependent.

3. Which is a subset of a linearly independent set of vectors is linearly independent.

4. Is linearly independent if and only if each of its subsets is linearly independent.

Proof. Left to the reader.

Example. To find the *least* number of linearly dependent vectors among

$$v_1 = \begin{bmatrix} 1 \\ -1 \\ 2 \\ -3 \end{bmatrix}, \ v_2 = \begin{bmatrix} 3 \\ -5 \\ 4 \\ -5 \end{bmatrix}, \ v_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \\ -1 \end{bmatrix}, \ v_4 = \begin{bmatrix} 4 \\ -7 \\ 5 \\ -6 \end{bmatrix}, \ v_5 = \begin{bmatrix} 2 \\ -3 \\ 3 \\ -4 \end{bmatrix}. \tag{5.18}$$

Solution of $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 + \alpha_5 v_5 = o$ yields $2\alpha_1 = \alpha_3 + \alpha_4 - \alpha_5$, $2\alpha_2 = -\alpha_3 - 3\alpha_4 - \alpha_5$ for arbitrary $\alpha_3, \alpha_4, \alpha_5$. For a nontrivial solution at most two of these can be set equal to zero, say $\alpha_4 = \alpha_5 = 0$. This leaves us with linearly dependent v_1, v_2, v_3, and the least number of linearly dependent vectors among v_1, v_2, v_3, v_4, v_5 is three.

Theorem 5.5. *Let $v_1, v_2, \ldots, v_m, v_{m+1}, \ldots, v_n$ be n linearly independent vectors. Vectors $v, v_{m+1}, \ldots, v_n$, where $v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m \neq o$ are linearly independent.*

Proof. We have that

$$\begin{aligned} \beta v + \beta_{m+1} v_{m+1} + \cdots + \beta_n v_n = \beta\alpha_1 v_1 + \beta\alpha_2 v_2 + \cdots + \beta\alpha_m v_m \\ + \beta_{m+1} v_{m+1} + \cdots + \beta_n v_n = o \end{aligned} \tag{5.19}$$

only if $\beta\alpha_1 = 0$, $\beta\alpha_2 = 0, \ldots, \beta\alpha_m = 0, \beta_{m+1} = 0, \ldots, \beta_n = 0$. At least one among $\alpha_1, \alpha_2, \ldots, \alpha_m$ must be different from zero, and necessarily $\beta = 0$. End of proof.

exercises

5.2.1. Are $v_1 = [3\ -5\ 1]^T$, $v_2 = [2\ -3\ 0]^T$, $v_3 = [-1\ 2\ -1]^T$ linearly independent?

5.2.2. Are $v_1 = [1\ 1\ 1\ 1]^T$, $v_2 = [1\ 1\ 0\ 0]^T$, $v_3 = [1\ 0\ 0\ 0]^T$ linearly independent?

5.2.3. Fix α so that

$$v_1 = \begin{bmatrix} 1 \\ -2 \\ \alpha \end{bmatrix},\ v_2 = \begin{bmatrix} -4 \\ -7 \\ 7 \end{bmatrix},\ v_3 = \begin{bmatrix} \alpha \\ 1 \\ -1 \end{bmatrix}$$

are linearly dependent.

5.2.4. Let a, b, c be linearly independent vectors. For what values of β are $a+\beta b, b+\beta c, c+\beta a$ linearly independent?

5.2.5. Let N be a nilpotent matrix of index m, $N^{m-1} \neq O, N^m = O$. Prove that if v is such that $N^{m-1}v \neq o$, then $v, Nv, N^2v, \ldots, N^{m-1}v$ are linearly independent.

5.2.6. Consider vectors $a = [a_1^T\ a_2^T]^T, b = [b_1^T\ b_2^T]^T, c = [c_1^T\ c_2^T]^T$. Show that if subvectors a_1, b_1, c_1 are linearly independent then so are a, b, c.

5.2.7 Let $v_1, v_2, \ldots, v_m$ be m linearly independent vectors in R^n. Prove that $v_2 - \alpha_2 v_1$, $v_3 - \alpha_3 v_1, \ldots, v_m - \alpha_m v_1$ are linearly independent.

5.2.8 Find the least number of linearly dependent vectors among

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \; v_2 = \begin{bmatrix} 3 \\ -1 \\ 2 \\ -2 \end{bmatrix}, \; v_3 = \begin{bmatrix} 2 \\ -2 \\ 1 \\ -1 \end{bmatrix}, \; v_4 = \begin{bmatrix} 4 \\ 0 \\ 3 \\ -3 \end{bmatrix}, \; v_5 = \begin{bmatrix} 5 \\ -3 \\ 3 \\ -3 \end{bmatrix}.$$

5.3 Bases and dimension

Linearly independent spans have special properties.

Definition. *A basis for vector space V is a linearly independent set of vectors which span V.*

A basis contains the *minimal* number of vectors that span V, or the *maximal* number of linearly independent vectors in V. In R^3 no more than one vector is needed to span a line, and no more than two vectors are needed to span a plane.

Example.

Set

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \; e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \; e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \; e_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{5.20}$$

is the *standard basis* for R^4, but

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \; v_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \; v_3 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \; v_4 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \tag{5.21}$$

is also a basis for R^4 (Verify!).

Theorem 5.6. *If $v_1, v_2, \ldots, v_n$ span vector space V, then any linearly independent set of vectors in V contains no more than n vectors.*

Proof. Assume for brevity that $n = 2$, and suppose that v_1, v_2 span V, and that v'_1, v'_2, v'_3 is a linearly independent set in V. Since v'_1, v'_2, v'_3 are in V there exist scalars A_{ij} so that

$$\begin{aligned} v'_1 &= A_{11}v_1 + A_{21}v_2 \\ v'_2 &= A_{12}v_1 + A_{22}v_2 \\ v'_3 &= A_{13}v_1 + A_{23}v_2. \end{aligned} \tag{5.22}$$

Completion of the proof rests on the demonstration that there are three scalars $\alpha_1, \alpha_2, \alpha_3$, some nonzero, such that $\alpha_1 v_1' + \alpha_2 v_2' + \alpha_3 v_3' = o$, contradicting the assumption on the linear independence of v_1', v_2', v_3'. Indeed,

$$\begin{aligned} \alpha_1 v_1' + \alpha_2 v_2' + \alpha_3 v_3' = (&\alpha_1 A_{11} + \alpha_2 A_{12} + \alpha_3 A_{13}) v_1 \\ &+ (\alpha_1 A_{21} + \alpha_2 A_{22} + \alpha_3 A_{23}) v_2 \end{aligned} \tag{5.23}$$

and there is a nontrivial choice of α's that renders the right-hand side of the equation zero since homogeneous system

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{5.24}$$

has more unknowns than equations and possesses therefore a nontrivial solution. End of proof.

Theorem 5.7. *Any two bases of vector space V have the same number of elements.*

Proof. Let $v_1, v_2, \ldots, v_m$ be one basis for V. By Theorem 5.6 every basis of V contains no more than m vectors. If $v_1', v_2', \ldots, v_n'$ is another basis for V, then $n \leq m$. By the reverse argument $m \leq n$, and hence $m = n$. End of proof.

Definition. *The number of vectors in any basis for V is the dimension,* $\dim(V)$*, of vector space V.*

Corollary 5.8. *If V is a vector space of dimension n,* $\dim(V) = n$*, then:*

1. Any set in V containing more than n vectors is linearly dependent.

2. No set in V containing fewer than n vectors can span V.

Proof.

1. Since $\dim(V) = n$, the space is spanned by n vectors and according to Theorem 5.6 cannot contain more than n linearly independent vectors.

2. Assumption of fewer than n vectors in the span of v would contradict the existence of a basis with n linearly independent vectors.

Corollary 5.9. *If V is a vector space of dimension n, then any n linearly independent vectors in V form a basis for V.*

Proof. Let $v_1, v_2, \ldots, v_n$ be n linearly independent vectors in V. Every vector $v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n$ is in V. No vector $v \in V$ exists that cannot be expressed by the linear combination, for otherwise it would imply the existence in V of more than n linearly independent vectors in contradiction to Theorem 5.6. End of proof.

Theorem 5.10. *Let $V : v_1, v_2, \ldots, v_n$. Removal of a linearly dependent vector from the span of V leaves the space unchanged.*

Proof. Say $n = 4$ with v_4 linearly depending on v_1, v_2, v_3,

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4,\ \ v_4 = \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3 \tag{5.25}$$

so that

$$v = (\alpha_1 + \beta_1 \alpha_4) v_1 + (\alpha_2 + \beta_2 \alpha_4) v_2 + (\alpha_3 + \beta_3 \alpha_4) v_3 = \alpha'_1 v_1 + \alpha'_2 v_2 + \alpha'_3 v_3 \tag{5.26}$$

with

$$\begin{aligned} \alpha_1 + \beta_1 \alpha_4 &= \alpha'_1 \\ \alpha_2 + \beta_2 \alpha_4 &= \alpha'_2 \\ \alpha_3 + \beta_3 \alpha_4 &= \alpha'_3. \end{aligned} \tag{5.27}$$

For any choice of $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, there are $\alpha'_1, \alpha'_2, \alpha'_3$, and conversely, for any choice of $\alpha'_1, \alpha'_2, \alpha'_3$, there are $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, with, possibly, $\alpha_4 = 0$. Hence v_1, v_2, v_3, v_4 and v_1, v_2, v_3 span the same vector space V. End of proof.

Example. To determine the dimension of vector space

$$V : v_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 2 \\ -1 \\ 2 \\ -1 \end{bmatrix}, v_3 = \begin{bmatrix} 5 \\ -1 \\ 5 \\ -1 \end{bmatrix}, v_4 = \begin{bmatrix} 3 \\ -2 \\ 3 \\ -2 \end{bmatrix}, \tag{5.28}$$

and find a basis for it we shall remove all linearly dependent vectors from the span of V. To this end we write the linear vector equation $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 = o$ in the tabular form

$$\begin{bmatrix} 1 & 2 & 5 & 3 \\ 0 & -1 & -1 & -2 \\ 1 & 2 & 5 & 3 \\ 0 & -1 & -1 & -2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = o \tag{5.29}$$

and determine that $\alpha_1 = -3\alpha_3 + \alpha_4, \alpha_2 = -\alpha_3 - 2\alpha_4$, implying that both v_3 and v_4 depend on v_1 and v_2. In fact, $v_3 = 3v_1 + v_2, v_4 = -v_1 + 2v_2$. Hence $V : v_1, v_2$, and $\dim(V) = 2$.

Theorem 5.11. *Let $v_1, v_2, \ldots, v_n$ be a set of n vectors in V. Then the following statements are equivalent:*

1. *Vectors $v_1, v_2, \ldots, v_n$ form a basis for V.*
2. *Every vector in V can be written as a unique linear combination of $v_1, v_2, \ldots, v_n$.*

Proof. Assume 1. Then $v_1, v_2, \ldots, v_n$ span V and any $v \in V$ is written as

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \tag{5.30}$$

or possibly

$$v = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n \tag{5.31}$$

so that

$$o = (\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \cdots + (\alpha_n - \beta_n)v_n. \tag{5.32}$$

Because $v_1, v_2, \ldots, v_n$ are linearly independent $\alpha_i - \beta_i = 0$ for all i.

Assume 2. Then

$$o = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \tag{5.33}$$

is certainly satisfied with $\alpha_1 = \alpha_2 = \cdots \alpha_n = 0$. But this choice of α's is unique, and hence $v_1, v_2, \ldots, v_n$ are linearly independent. End of proof.

Lemma 5.12. *Matrix $A(m \times n)$, $m \geq n$, possesses a left inverse if and only if the dimension of its column space C is n, that is, if and only if its columns are linearly independent.*

Proof. According to Theorem 2.12 a left inverse to $A = [a_1 a_2 \ldots a_n]$ exists if and only if A is of full column rank. Matrix A is of full column rank only if no general elementary operation exists that annuls one of its columns, meaning that $\alpha_1 a_1 + \alpha_2 a_2 + \cdots + \alpha_n a_n = o$ only if $\alpha_1 = \alpha_2 = \cdots \alpha_n = 0$. End of proof.

For the particular case of a square matrix, Lemma 5.12 states that $A = A(n \times n)$ is nonsingular if and only if its column space is of dimension n.

Theorem 5.13. *If $W : w_1, w_2, \ldots, w_m$ of dimension $m, dim(W) = m$, is a subspace of $V : v_1, v_2, \ldots, v_n$ of dimension $n, dim(V) = n,\ n \geq m$, then exactly m vectors in the basis for V can be replaced by the m vectors of the basis for W. In other words, if $W \subset V$, then the respective bases for the vector spaces may be chosen, generically, so that $W : v_1, v_2, \ldots, v_m$, and $V : v_1, v_2, \ldots, v_m, v_{m+1}, \ldots, v_n$.*

Proof. Say $W\colon w_1, w_2$, and $V\colon v_1, v_2, v_3, v_4$. Since w_1 and w_2 are both in V they may be expressed by the linear combinations

$$\begin{aligned} w_1 &= A_{11}v_1 + A_{21}v_2 + A_{31}v_3 + A_{41}v_4 \\ w_2 &= A_{12}v_1 + A_{22}v_2 + A_{32}v_3 + A_{42}v_4. \end{aligned} \tag{5.34}$$

Linear independence stipulates that

$$\begin{aligned} \beta_1 w_1 + \beta_2 w_2 = (\beta_1 A_{11} + \beta_2 A_{12})v_1 + (\beta_1 A_{21} + \beta_2 A_{22})v_2 \\ + (\beta_1 A_{31} + \beta_2 A_{32})v_3 + (\beta_1 A_{41} + \beta_2 A_{42})v_4 = o \end{aligned} \tag{5.35}$$

only if $\beta_1 = \beta_2 = 0$ and

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \\ A_{41} & A_{42} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{5.36}$$

implying that matrix A in eq.(5.36) is of full column rank. Say that as a result of a sequence of elementary *column* operations

$$A \rightarrow \begin{bmatrix} & \\ 1 & \\ \times & \\ & 1 \end{bmatrix}. \tag{5.37}$$

Then w_1, w_2, v_1, v_3 are linearly independent. Indeed, $\beta_1 w_1 + \beta_2 w_2 + \beta_3 v_1 + \beta_4 v_3 = o$, together with eq.(5.34), require that

$$\begin{bmatrix} A_{11} & A_{12} & 1 & \\ A_{21} & A_{22} & & \\ A_{31} & A_{32} & & 1 \\ A_{41} & A_{42} & & \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = o. \tag{5.38}$$

In view of eq.(5.37) the matrix in system (5.38) is of full column rank and $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ only. End of proof.

Definition. *Let S_1 and S_2 be two sets of vectors in R considered to be spans for subspaces V_1 and V_2, respectively of R. Then $V : S_1 \cup S_2$ is a subspace V of R that is the union of V_1 and V_2, $V = V_1 \cup V_2$.*

Notice that $V_1 \cup V_2$ is not the set of all vectors in V_1 and V_2. Say $V_1 : e_1, V_2 : e_2, V_3 : e_3$. Union $V_1 \cup V_2 \cup V_3 : e_1, e_2, e_3$ is R^3, but $v = [1\ 1\ 1]^T$ is neither in V_1, nor in V_2, nor in V_3.

Lemma 5.14. *Subspaces V_1 and V_2 of R are disjoint, i.e. they share no common nonzero element, if and only if sets S_1 and S_2 that contain their respective bases are such that the collective set $S = S_1 \cup S_2$ is linearly independent. Then $\dim(V) = \dim(V_1 \cup V_2) = \dim(V_1) + \dim(V_2)$.*

Proof. Suppose that set $S_1 \cup S_2$ is linearly independent and say v is an element shared by both V_1 and V_2 so that

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n \ \text{ and } \ v = \beta_1 w_1 + \beta_2 w_2 + \ldots + \beta_m w_m. \tag{5.39}$$

Then

$$o = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n - \beta_1 w_1 - \beta_2 w_2 - \ldots - \beta_m w_m \tag{5.40}$$

and since S_1 and S_2 are linearly independent $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$, and $\beta_1 = \beta_2 = \ldots = \beta_m = 0$. The only element common to both V_1 and V_2 is the zero vector.

Suppose that S_1 and S_2 are linearly dependent. Then

$$o = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n + \beta_1 w_1 + \beta_2 w_2 + \ldots + \beta_m w_m \tag{5.41}$$

holds with at least one nonzero α_k and one nonzero β_k. There is now a nonzero element

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n = -\beta_1 w_1 - \beta_2 w_2 - \ldots - \beta_m w_m \tag{5.42}$$

that is in both V_1 and V_2.

Set $S = S_1 \cup S_2$ is a basis for $V = V_1 \cup V_2$ and $\dim(V) = m + n$. End of proof.

Definition. *Let S_1, S_2, S_3 be three sets of vectors in R so that $S = S_1 \cup S_2 \cup S_3$ is linearly independent. If $V_1 : S_1 \cup S_2$ and $V_2 : S_2 \cup S_3$, then $V_3 : S_2$ is their intersection, $V_3 = V_1 \cap V_2$.*

Set S_2 holds the basis for V_3, and if S_2 includes k vectors, then $\dim(V_3) = \dim(V_1 \cap V_2) = k$. The only vectors shared by V_1 and V_2 are those of V_3.

Theorem 5.15. *Let V_1 and V_2 be two vector subspaces of R. Then*

$$\dim(V_1) + \dim(V_2) = \dim(V_1 \cup V_2) + \dim(V_1 \cap V_2). \tag{5.43}$$

Proof. Let S_1, S_2, S_3 be three sets of vectors in R so that $S_1 \cup S_2 \cup S_3$ is linearly independent. Let S_1 contain m vectors, S_2 contain k vectors, and S_3 contain n vectors. If $S_1 \cup S_2$ is a basis for V_1 and $S_2 \cup S_3$ is a basis for V_2, then $\dim(V_1) = m+k$, $\dim(V_2) = n+k$, $\dim(V_1 \cup V_2) = m+n+k$, $\dim(V_1 \cap V_2) = k$, and hence the equation. End of proof.

For instance, if $S_1 : v_1, v_2, v_3;\ S_2 : v_4, v_5;\ S_3 : v_6, v_7;\ V_1 : S_1 \cup S_2,\ V_2 : S_2 \cup S_3$; then $\dim(V_1) = 5$, $\dim(V_2) = 4$, $\dim(V_1 \cup V_2) = 7$. $\dim(V_1 \cap V_2) = 2$, and $7 + 2 = 5 + 4$.

Example. To find the intersection of

$$V : \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} \text{ and } W : \begin{bmatrix} 2 \\ 1 \\ 2 \\ -5 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ -1 \\ -1 \end{bmatrix}. \tag{5.44}$$

Write the span of V as v_1, v_2, v_3, and that of W as w_1, w_2. The intersection of V and W consists of all vectors such that

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = -\beta_1 w_1 - \beta_2 w_2 \tag{5.45}$$

or in matrix vector form

$$\begin{bmatrix} 1 & 1 & 1 & 2 & -1 \\ 1 & -1 & 0 & 1 & 3 \\ 1 & 1 & -1 & 2 & -1 \\ 1 & -1 & 0 & -5 & -1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5.46}$$

We solve the system for $\alpha_1 = 0,\ \alpha_3 = 0,\ \alpha_2 = 7/3\beta_2,\ \beta_1 = -2/3\beta_2$ for arbitrary β_2, and $v = 7/3\beta_2 v_2 = 2/3\beta_2 w_1 - \beta_2 w_2$. Span $V \cap W$ is $v_2 = [1\ \ -1\ 1\ \ -1]^T$ and its dimension is one.

The language of vector spaces, the high language of modern mathematics, is indicative rather than constructive. Such statements as: *linear system $Ax = f$ has a solution if and only if vector f is in the column space of A*, is concise and intuitive, but requires lengthy computational interpretations before it is translated into a practical criterion.

exercises

5.3.1. Is the set of all vectors $x = [x_1\ x_2\ x_3]^T$ subject to the condition $x^T a = 0,\ a = [1\ -1\ 1]^T$, a vector space? If yes, write a span for it, and determine its dimension.

5.3.2. Write all vectors in R^4 that are not in

$$V\colon \begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\end{bmatrix}.$$

5.3.3. For

$$V: \begin{bmatrix}3\\1\\3\\1\end{bmatrix}, \begin{bmatrix}1\\-1\\1\\-1\end{bmatrix} \quad \text{and} \quad W: \begin{bmatrix}5\\-2\\-1\\-2\end{bmatrix}, \begin{bmatrix}2\\-1\\0\\-1\end{bmatrix}$$

write a basis for $V \cap W$.

5.4 Linear transformations

Linear transformations have most significant geometrical interpretations, and we have considered them already as such in Chapter 4. Here we want to reconsider the terminology of linear transformations in the algebraic context of higher-dimensional vector spaces.

Consider, then, the linear transformation $x' = Ax, A = A(m \times n) = [a_1\ a_2, \ldots, a_n]$, in which x is a *variable* vector selected at will from *domain* R^n of the transformation. Variable vector x' is the *image* of x under the transformation. It is confined to the column space of $A,\ x' = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$, a subspace of R^m that is the *range* of the transformation.

The linear transformation, or mapping

$$\begin{bmatrix}x'_1\\x'_2\\x'_3\end{bmatrix} = \begin{bmatrix}1 & 2 & 1\\-2 & -1 & 1\\1 & 2 & 1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = x_1\begin{bmatrix}1\\-2\\1\end{bmatrix} + x_2\begin{bmatrix}2\\-1\\2\end{bmatrix} + x_3\begin{bmatrix}1\\1\\1\end{bmatrix},\ x' = Ax, \tag{5.47}$$

for example, takes arbitrary vector $x \in R^3$ and dispatches it to column space $C(A)$ of matrix A. We symbolically articulate this algebraic transaction by writing the compact message $T : R^3 \rightarrow C$. We numerically verify that the columns of A are linearly dependent so that $\dim(C) = 2$, and conclude that three-dimensional vector space R^3 is mapped by linear transformation (5.47) into two-dimensional vector subspace C of R^3. Linear transformation (5.47) assigns to every vector $x \in R^3$ a unique image vector x', but not conversely. Vector $x' = [4\ \ -2\ 4]^T$, for instance, is the image of $x = [1\ 1\ 1]$, of $x = [2\ 0\ 2]^T$, and, in fact, of $x = [\alpha\ 2 - \alpha\ \alpha]$ for any α. All this comes from basic facts of the solution of systems of linear equations.

By contrast, linear transformation

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -2 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{5.48}$$

which is with a nonsingular, or invertible, matrix A maps every $x \in R^3$ into a distinct image $x' \in R^3$, and *conversely*, since not only is $x' = Ax$ but also reversely $x = A^{-1}x'$. Linear transformation (5.48) is a one-to-one transformation that maps the whole of R^3 *onto* itself.

We may find it convenient or expedient to span R^n with bases other than the standard. Say that instead of using the standard basis for R^3 we chose to express x as

$$x = \alpha_1 \begin{bmatrix} 1 \\ \\ \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ \\ \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \tag{5.49}$$

in which variables $\alpha_1, \alpha_2, \alpha_3$ are sometimes called the *components* or *coordinates* of x in the chosen basis. On the other hand, every vector x' in the column space C of matrix A affecting linear transformation (5.47) may be written as

$$x' = \beta_1 \begin{bmatrix} 1 \\ \\ 1 \end{bmatrix} + \beta_2 \begin{bmatrix} \\ 1 \\ \\ \end{bmatrix} = \begin{bmatrix} 1 & \\ & 1 \\ 1 & \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \tag{5.50}$$

for arbitrary β_1, β_2. With eqs. (5.49) and (5.50) transformation (5.47) becomes

$$\begin{bmatrix} 1 & \\ & 1 \\ 1 & \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \tag{5.51}$$

assuming finally the form

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 4 \\ -2 & -3 & -2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \tag{5.52}$$

so that if $\alpha_1 = 1, \alpha_2 = -2, \alpha_3 = 3$, then $\beta_1 = 7, \beta_2 = -2$, and $x = [2\ 1\ 3]^T, x' = [7\ \ -2\ 7]^T$, which we could have gotten also from eq. (5.47).

One more example. Consider the two vector spaces

$$V : v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \text{ and } W : w_1 = \begin{bmatrix} 1 \\ \\ -1 \\ \\ \end{bmatrix}, w_2 = \begin{bmatrix} \\ 1 \\ \\ -1 \end{bmatrix} \tag{5.53}$$

that are a pair of disjoint two-dimensional subspaces of R^4. Any $v \in V$ is written by its components as $v = \alpha_1 v_1 + \alpha_2 v_2$, and any $w \in W$ is written in terms of its components as $w = \beta_1 w_1 + \beta_2 w_2$. A one-to-one linear transformation between spaces V and W is induced by a one-to-one linear relationship between the components of $v \in V$ and $w \in W$. For instance,

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \tag{5.54}$$

which is essentially a one-to-one mapping of R^2 onto itself. The fact the one-to-one linear transformation $T : V \to W$ reverts to a one-to-one linear transformation of R^2 onto itself is expressed by saying that V and W are *isomorphic* to R^2. Any two-dimensional vector space is, in fact, isomorphic to R^2, any three-dimensional vector space is isomorphic to R^3, and so on. Geometrically speaking, R^2 is a plane, and so are V and W in the generalized sense of higher dimensions. In this sense, not only are V and W isomorphic to R^2, but they are also isomorphic to each other, and any one-to-one linear transformation between them is said to be an isomorphism.

Since a linear transformation of two vector spaces is produced by a matrix we shall refer concretely to matrices as they affect the transformation rather than to the transformation itself.

Theorem 5.16. *Let $v_1, v_2, \ldots, v_n$ be a basis for the n dimensional vector space V, and let $w_1, w_2, \ldots, w_n$ be n vectors in vector space W. Then there is one and only one linear transformation T from V into W so that $Tv_i = w_i$ $i = 1, 2, \ldots, n$.*

Proof. Let v_j be the jth column of matrix V and w_j the jth column of matrix W, and write $TV = W$. By Lemma 5.12 V is nonsingular and $T = V^{-1}W$. To prove uniqueness assume also that $T'V = W$. This means that $(T - T')V = O$, and since V is invertible $T - T' = O$. End of proof.

5.5 Row and column spaces

The rows and columns of matrix $A = A(m \times n)$, considered as vectors in R^n and R^m, possess some remarkable interdependent properties.

Theorem 5.17. *Let T and T' be two nonsingular matrices. Then A and TA have the same row space, and A and AT' have the same column space.*

Proof. Because $B = AT'$ and $A = BT'^{-1}$ the columns of B are in the column space of A, and the columns of A in the column space of B. Matrices A and $B = AT'$ have the same column space. A similar argument holds for the rows. End of proof.

Theorem 5.17 effectively states that the two vector spaces

$$V : v_1, v_2 \text{ and } W : w_1 = T_{11}v_1 + T_{21}v_2, w_2 = T_{12}v_1 + T_{22}v_2 \tag{5.55}$$

are equal if matrix T has an inverse. That $W \subset V$ is obvious by the linear combination. By virtue of the invertability of matrix T, matrix $T' = T^{-1}$ also exists so that conversely $v_1 = T'_{11}w_1 + T'_{21}w_2, v_2 = T'_{12}w_1 + T'_{22}w_2$, which we readily verify by substitution.

Corollary 5.18. *Row equivalent matrices have the same row space, and column equivalent matrices have the same column space.*

Proof. In Theorem 5.17 let T and T' be two elementary operation matrices. End of proof.

Determination of bases for the row and column spaces of a matrix requires the reduction of the matrix by means of elementary operations to row and column echelon form. A typical *column* echelon matrix is written below:

$$A = \begin{bmatrix} 1 & & \\ \times & & \\ \times & & \\ \times & 1 & \\ \times & \times & 1 \\ \times & \times & \times \end{bmatrix}. \tag{5.56}$$

Theorem 5.19. *Let A be a column (row) echelon matrix. The nonzero column (row) vectors of A hold a basis for the column (row) space of A.*

Proof. Obviously the nonzero column vectors of A are linearly independent. End of proof.

Example.

$$\alpha_1 \begin{bmatrix} 1 \\ \times \\ \times \\ \times \\ \times \\ \times \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ \times \\ \times \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ \times \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{5.57}$$

only if $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

Theorem 5.17 and Corollary 5.18 assure us that elementary row operations do not change the row space of matrix A, and similarly elementary column operations do not change the column space of A. Theorem 5.19 says that for matrix A in row echelon form a basis for the row space is written in the nonzero rows of A, and the same holds true for the columns. Hence the way to write bases for the row space and column space of any matrix.

Example. To find bases for the row and column spaces of A. By elementary column operations

$$A = \begin{bmatrix} 1 & 1 & -1 & 2 & 2 \\ -1 & 1 & 0 & 3 & 2 \\ 2 & 1 & 1 & 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & & & & \\ -1 & 2 & -1 & 5 & 4 \\ 2 & -1 & 3 & -3 & -3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & & & & \\ -1 & 2 & & & \\ 2 & -1 & 5/2 & -1/2 & -1 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & & & & \\ -1 & 2 & & & \\ 2 & -1 & 5 & & \end{bmatrix} \rightarrow \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \end{bmatrix} \tag{5.58}$$

we find basis e_1, e_2, e_3 for the column space C of A, and $\dim(C) = 3$. By elementary row operations

$$A = \begin{bmatrix} 1 & 1 & -1 & 2 & 2 \\ -1 & 1 & 0 & 3 & 2 \\ 2 & 1 & 1 & 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & -1 & 2 & 2 \\ & 2 & -1 & 5 & 4 \\ & -1 & 3 & -3 & -3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & -1 & 2 & 2 \\ & 2 & -1 & 5 & 4 \\ & & 5/2 & -1/2 & -1 \end{bmatrix}$$
$$\rightarrow \begin{bmatrix} 1 & 1 & & 9/5 & 8/5 \\ & 2 & & 24/5 & 18/5 \\ & & 5/2 & -1/2 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & & & -3/5 & -1/5 \\ & 2 & & 24/5 & 18/5 \\ & & 5/2 & -1/2 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 5 & & & -3 & -1 \\ & 5 & & 12 & 9 \\ & & 5 & -1 & -2 \end{bmatrix} \tag{5.59}$$

and the row space R of A is

$$R: \begin{bmatrix} 5 \\ 0 \\ 0 \\ -3 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ 0 \\ 12 \\ 9 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 5 \\ -1 \\ -2 \end{bmatrix} \tag{5.60}$$

and $\dim(R) = 3$.

exercises

5.5.1. Show that the set of all solutions x to the homogeneous system

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} x = o$$

constitutes a vector space. Write a basis for it.

5.6 Rank and nullity

Linearly independent vectors are mapped in a nonsingular linear transformation into linearly independent vectors. This is a basic fact of nonsingular mapping and it has interesting consequences.

Lemma 5.20. *Let $v_1, v_2, \ldots, v_n$ of space V be linearly independent, and A nonsingular. Then $Av_1, Av_2, \ldots, Av_n$ are linearly independent.*

Proof. If $v_1, v_2, \ldots, v_n$ are linearly independent then

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = v \neq o \tag{5.61}$$

for any nontrivial choice of $\alpha_1, \alpha_2, \ldots, \alpha_n$. Since A is nonsingular $Av \neq o$ if $v \neq o$ and

$$\alpha_1 A v_1 + \alpha_2 A v_2 + \cdots + \alpha_n A v_n = Av \neq o. \tag{5.62}$$

End of proof.

Theorem 5.21. *Let T and T' be nonsingular matrices. Then A and TA have a column space of the same dimension, and A and AT' have a row space of the same dimension.*

Proof. If the columns of A are $a_1, a_2, \ldots, a_n$, then the columns of TA are $Ta_1, Ta_2, \ldots, Ta_n$. Since linearly independent vectors are mapped into linearly independent vectors by the nonsingular T the number of linearly independent columns is the same for A as for TA. The same argument holds for A^T and ${T'}^T A^T$. End of proof.

Theorem 5.22. *The totality of solutions of $Ax = o$ constitute a vector space.*

Proof. Let $v_1, v_2, \ldots, v_k$ be an entire set of linearly independent solutions to $Ax = o$. Then $x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k$ is such that $Ax = o$ for any $\alpha_1, \alpha_2, \ldots, \alpha_k$. No $v, Av = o$ exists that cannot be written as $v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k$ for this would imply that $v_1, v_2, \ldots, v_k, v$ are linearly independent contrary to the assumption on the vectors. End of proof.

Definition. *The space of all vectors x such that $Ax = o$ is the nullspace, or kernel (sometimes written* $\ker(A)$*) of A. The nullity of A is the dimension of its nullspace, and the rank of A is the dimension of its column space.*

Theorem 5.23. *Let $A = A(m \times n)$. The dimension of the column space of A (column rank (A)), and the dimension of the row space of A (row rank (A)), are equal:*

$$\textit{row rank } (A) = \textit{ column rank } (A) = \textit{rank } (A). \tag{5.63}$$

Proof. According to Theorems 5.17 and 5.21 elementary row and column operations leave the dimensions of the row and column spaces of A unchanged. Elementary row operations reduce A to a row echelon form, then elementary column operations bring it to the form

$$A' = \begin{bmatrix} I(r \times r) & O \\ O & O \end{bmatrix} \tag{5.64}$$

that shows the equality. End of proof.

Theorem 5.23 is a restatement of Theorem 1.16.

Theorem 5.24. *If $A = A(m \times n)$, then*

$$rank(A) + nullity(A) = n. \tag{5.65}$$

Proof. Let $v_1, v_2, \ldots, v_k,\ v_{k+1}, \ldots, v_n$ be a basis for domain R^n of $Ax = x'$, such that $v_1, v_2, \ldots, v_k$ is a basis for nullspace N of $A, Av_i = o\ i = 1, 2, \ldots, k$. We shall prove that $Av_{k+1}, Av_{k+2}, \ldots, Av_n$ is a basis for range R, or column space C, of A.

Indeed, if

$$x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k + \alpha_{k+1} v_{k+1} + \cdots + \alpha_n v_n \tag{5.66}$$

then

$$x' = Ax = o + \alpha_{k+1} Av_{k+1} + \cdots + \alpha_n Av_n \tag{5.67}$$

and $Av_{k+1}, Av_{k+2}, \ldots, Av_n$ span R. These vectors are also linearly independent. To show this we write

$$\beta_{k+1} Av_{k+1} + \cdots + \beta_n Av_n = A(\beta_{k+1} v_{k+1} + \cdots + \beta_n v_n) = o \tag{5.68}$$

implying that either $\beta_{k+1} v_{k+1} + \cdots + \beta_n v_n$ is in the nullspace of A or the β's are all zero. But the vector sum is not in N and hence $\beta_{k+1} = \beta_{k+2} = \cdots = \beta_n = 0$. Now $\dim(N) = k,\ \dim(R) = n - k$, and $\dim(N) + \dim(R) = n$. End of proof.

Notice how the proof to this theorem is done in the spirit of abstract linear algebra without dipping into the tabular nature and structure of matrix A or vector x. Otherwise we may argue that if the rank of system $Ax = o$ is r, then solution vector x includes $n - r$ arbitrary components making for a nullspace of that dimension , and $r + (n - r) = n$.

The nullity of an $n \times n$ nonsingular matrix is zero and the dimension of both its column space and row space is n.

Rank(A) of $A = A(m \times n)$ counts the largest number of linearly independent columns (rows) in matrix A, while nullity(A) counts the largest number of linearly independent vectors $x \in R^n$ for which $Ax = o$.

Example. To find a basis for the nullspace of A:

$$Ax = \begin{bmatrix} 1 & -1 & -2 \\ 2 & -1 & -3 \\ -1 & 2 & 3 \\ -3 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \; EAx = \begin{bmatrix} 1 & & -1 \\ & 1 & 1 \\ & & \\ & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5.69}$$

Hence $x_1 = x_3,\; x_2 = -x_3$, and $n = [1 \;\, -1 \;\, 1]^T$, is a basis for the one-dimensional nullspace.

Theorem 5.25.

$$\text{nullity}(AB) \geq \text{nullity}(B) \tag{5.70}$$

$$\text{rank}(AB) \leq \text{rank}(A), \quad \text{rank}(AB) \leq \text{rank}(B). \tag{5.71}$$

Proof. Obviously for every vector x such that $Bx = 0$, also $A(Bx) = 0$, and hence the nullspace of B is contained in that of AB.

Say $A = A(m \times k), B = B(k \times n)$ so that $AB = AB(m \times n)$. Then by Theorem 5.24 rank(AB) $= n-$nullity(AB), and rank(B) $= n-$nullity(B), and the first rank inequality of the theorem results from the inequality of the nullities. The second rank inequality is obtained from rank(AB) = rank$((AB)^T)$ = rank$(B^T A^T)$ $\leq$ rank(A^T) = rank(A). End of proof.

Theorem (the Frobenius rank inequality) 5.26.

$$\text{nullity}(AB) + \text{nullity}(BC) \geq \text{nullity}(ABC) + \text{nullity}(B) \tag{5.72}$$

$$\text{rank}(AB) + \text{rank}(BC) \leq \text{rank}(ABC) + \text{rank}(B). \tag{5.73}$$

Proof. Let $x_1, x_2, \ldots, x_k$ be a basis for $N(BC)$, and $x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_{k+1+l}$ a basis for $N(ABC)$, the nullspace of ABC, so that nullity$(ABC)-$nullity$(BC) = l + 1$.

Vectors $(BC)x_{k+1}, \ldots, (BC)x_{k+1+l}$ are linearly independent. Indeed,

$$\alpha_{k+1}(BC)x_{k+1} + \cdots + \alpha_{k+1+l}(BC)x_{k+1+l} = (BC)(\alpha_{k+1}x_{k+1} + \cdots + \alpha_{k+1+l}x_{k+1+l}) = o \tag{5.74}$$

happens only if $\alpha_{k+1}x_{k+1} + \cdots + \alpha_{k+1+l}x_{k+1+l} = o$ because vectors $x_{k+1}, \cdots, x_{k+1+l}$ are not in the nullspace of BC, and it results, by virtue of the fact that $x_{k+1}, \cdots, x_{k+1+l}$ are themselves linearly independent, that $\alpha_{k+1} = \cdots = \alpha_{k+1+l} = 0$.

Vectors $Cx_{k+1}, \cdots, Cx_{k+1+l}$ are also linearly independent since

$$\alpha_{k+1}Cx_{k+1} + \cdots + \alpha_{k+1+l}Cx_{k+1+l} = C(\alpha_{k+1}x_{k+1} + \cdots + \alpha_{k+1+l}x_{k+1+l}) = o \tag{5.75}$$

happens only if $\alpha_{k+1}x_{k+1} + \cdots + \alpha_{k+1+l}x_{k+1+l} = o$, because vectors $x_{k+1}, \cdots, x_{k+1+l}$ are not in the nullspace of C. The $l+1$ linearly independent vectors $Cx_{k+1}, \cdots, Cx_{k+1+l}$ are thus in the nullspace of $AB, (AB)(Cx_{k+1}) = \cdots = (AB)(Cx_{k+1+l}) = o$, but not in the nullspace of $B, B(Cx_{k+1}) \neq o \cdots B(Cx_{k+1+l}) \neq o$. Hence nullity$(AB)-$nullity$(B) \geq l+1$, and the inequality on the nullities of the theorem results.

Say $A = A(m \times k), B = B(k \times l), C = C(l \times n)$ so that $AB = AB(m \times l), BC = BC(k \times n), ABC = ABC(m \times n)$. Then by Theorem 5.24 nullity$(AB) = l-$rank(AB), nullity$(B) = l-$rank(B), nullity$(BC) = n-$rank(BC), nullity$(ABC) = n-$rank(ABC), and the rank inequality of the theorem follows. End of proof.

exercises

5.6.1. Let solution x to $Ax = o$, $A = A(m \times n)$, be such that $x_1, x_2, \ldots, x_k$ are independent and $x_{k+1}, \ldots, x_n$ are dependent. Show that $x = x_1v_1 + x_2v_2 + \ldots + x_kv_k$ with linearly independent $v_1, v_2, \ldots, v_k$.

5.6.2. Write bases for the column space, row space, and nullspace of

$$A = \begin{bmatrix} 5 & -1 & 3 & 4 & -7 \\ 4 & 0 & 1 & 3 & -5 \\ 3 & 1 & -1 & 2 & -3 \end{bmatrix}.$$

Also for the nullspace of A^T.

5.6.3. Does the set of all solutions to

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} x = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

constitute a vector space?

5.6.4. Show that the nullspace $N(A)$ of

$$A = \begin{bmatrix} B \\ B' \end{bmatrix}$$

is $N(B) \cap N(B')$.

5.6.5. Consider the three matrices

$$M_1 = \begin{bmatrix} 4 & -5 \\ 11 & 1 \end{bmatrix}, \ M_2 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \ M = \begin{bmatrix} 4 & -5 \\ 11 & 1 \end{bmatrix}.$$

Are M_1 and M_2 linearly independent? Is M in the vector space spanned by M_1, M_2?

5.6.6. Is the set of all 2×2 symmetric matrices a vector space? If yes, write a basis for the space and count its dimension.

5.6.7. What is $U \cap L$ if U is the vector space of all $n \times n$ upper-triangular matrices and L is the vector space of all $n \times n$ lower-triangular matrices?

5.6.8. What is $S \cap S'$ if S' is the vector space of all symmetric matrices and S' the vector space of all skew symmetric matrices?

5.6.9. Show that the totality of matrices that commute with

$$A = \begin{bmatrix} 1 & & 1 \\ & -1 & \\ & & 1 \end{bmatrix}$$

constitute a vector space. Write a basis for it.

5.6.10. Does the totality of the right inverses to

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

constitute a vector space?

5.6.11. Show that the totality of circulant matrices

$$C = \begin{bmatrix} \alpha & \beta & \gamma \\ \gamma & \alpha & \beta \\ \beta & \gamma & \alpha \end{bmatrix}$$

for arbitrary α, β, γ constitute a vector space. Find a basis for the space and determine its dimension. Generalize the result to $n \times n$ circulant matrices.

5.6.12. Show that the mapping $Y(X) = AX + XB$ is linear,

$$Y(\alpha X + \alpha' X') = \alpha Y(X) + \alpha' Y(X').$$

Find the range and the nullspace of the transformation for

$$A = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}, B = \begin{bmatrix} -1 & \\ & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & \\ & 2 \end{bmatrix}, B = \begin{bmatrix} 2 & \\ & 1 \end{bmatrix}.$$

5.6.13. Is the set of all upper-triangular matrices U such that $U_{ii} = 1$ a vector space?

5.6.14. Is the linear combination of two $n \times n$ permutation matrices a permutation matrix?

5.6.15. If the set of all matrices that commute with

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$$

constitutes a vector space, then write a basis for it and determine its dimension.

5.6.16. Count the number of linearly independent 2×2 matrices of rank 1.

5.6.17. Are $\phi_1 = 1,\ \phi_2 = \xi,\ \phi_3 = \xi^2$ linearly independent?

5.7 Analogies and abstractions

Linear algebra is the mathematics of linear combinations and transformations and as such bids generalizations to mathematical objects other than ordered lists of numbers. Indeed, our introduction to vector spaces as put forth in eq.(5.1) does not explicitly stipulate that $v_1, v_2, \ldots, v_n$, the n vectors that span the space, be vectors in the traditional sense. All it actually requires is that we be able to form linear combinations of these objects with addition and scalar multiplication as we do for numbers. We sense here a great opportunity for abstraction, but must beware not to over carry the abstraction into obfuscation. In our willingness to speak in generalities about the linear combination of objects, elements and things (i.e. tangibles) we risk creating an illusion. Objects or "things" cannot be added or multiplied, *only numbers can be added and multiplied*, and the linear combination of "elements" remains meaningless until and unless the elements have revealed their numerical faces. In writing $\phi_1(\xi)+\phi_2(\xi)$ we do not intend the addition of the two mathematical objects that we call functions, but rather the addition of the numerical values of ϕ_1 and ϕ_2 for some value of variable ξ.

Yet we may consider the pattern and structure of vector spaces without giving a thought to the nature of vectors, and still have most of the subject intact and in place. This is because the prevalence of vector spaces is metaphorical. Few of the theorems of the previous sections netted us anything computationally new or theoretically revealing. In fact, the whole edifice of vector spaces as we have presented it rests on the theorem that a homogeneous system of m equations in n unknowns has a nontrivial solution if $n > m$. The concrete answer to the basic question as to whether or not a given vector is "in" a certain vector space reverts, for vectors in R^n, to the question of whether or not a linear system of equations is consistent. In the same way the proofs of vector space theorems consist in large part in translating the new language into meaningful statements on linear systems of equations. Saying that the rows of a matrix are linearly independent is an elegant, concise, idiomatic statement to the effect that elementary row operations done on the matrix will never annul a row.

The basic concepts of vector space theory such as linear independence, span, basis, and dimension are also not predicated on the specific nature of what we may call vectors. Examples will make this clear.

Take for instance matrices. The (infinite!) set of all $M = M(3 \times 2)$ matrices created by the linear combination

$$M = \alpha_1 \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 & 1 \\ -2 & 0 \\ 1 & -1 \end{bmatrix} = \alpha_1 M_1 + \alpha_2 M_2 \tag{5.76}$$

with arbitrary real scalar variables α_1, α_2 constitutes a vector space, say M, spanned by M_1, M_2, which is a subspace of $R^{3\times2}$ —the space of all 3×2 real matrices.

It should be perfectly clear to us what is meant by linearly independent matrices, and a vector space of matrix elements has a span, a basis, and a dimension. Matrices M_1 and M_2 are linearly independent, they are a basis for space M which is two dimensional.

Or we may speak of function vector spaces. Let $F : \phi(\xi),\ \phi_2(\xi), \ldots, \phi_n(\xi)$ be a set of n functions defined in the interval $0 \leq \xi \leq 1$. Linear combination

$$\phi = \alpha_1\phi_1 + \alpha_2\phi_2 + \cdots + \alpha_n\phi_n \tag{5.77}$$

for real α's is defined for all ξ in that interval, and the totality of ϕ's thus created constitutes a vector space, say Φ. Functions $\phi_1(\xi),\ \phi_2(\xi), \ldots, \phi_n(\xi)$, span the space, and if they are linearly independent, then $\dim\Phi = n$. For instance, $\phi_1 = \xi, \phi_2 = \xi^2$ are linearly independent, never can $\alpha_1\phi_1 + \alpha_2\phi_2 = 0$ with a nonzero α. To show this we select nonzero $\xi = \xi_1$, and $\xi = -\xi_1$, and have that $\alpha_1\xi_1 + \alpha_2\xi_1^2 = 0$, and $-\alpha_1\xi_1 + \alpha_2\xi_1^2 = 0$, from which it readily results that, necessarily, $\alpha_1 = \alpha_2 = 0$.

Functions, however, are of a mathematical nature vastly more complicated than lists of numbers, and their use raises delicate questions on continuity, differentiability, and integrability. Which of these properties, we wonder, does the span pass on to the space? Here we must resort to analysis for help, and it assures us, for instance, that if the span is continuous on some interval, then every function of the space is continuous on that interval. But no finite dimensional vector space of continuous basis functions can include every continuous function. By a stretch of the imagination and by dint of a careful limiting process we may consider the collection of all functions continuous on an interval as comprising a vector space of *infinite* dimensions with a never-ending set of appropriate basis functions. This space is an ideal; in practice we must limit ourselves to finite dimensional vector spaces of functions, albeit of an immense size. There are ways, such as by using finite elements, for devising finite dimensional function vector spaces that hold a good *approximation* to any continuous function on a certain interval. All this is mathematically fascinating and is of great practical value, but belongs in the distinct and distinguished disciplines of functional analysis and function approximations.

And if the vector is abstracted, then so must be the matrix vector multiplication. Abstract linear algebra dissociates itself from matrix algebra by splitting the nature of the matrix. On the one hand it retains it, with its algebra, as a two-dimensional array of numbers, for its essential role in systems of equations and linear transformations. On the other hand the matrix vector multiplication $w = Av$ is abstracted into a general linear operation on general vector v to produce general vector w.

With functions the linear operation may be differentiation. If operator D is such that $D(\xi^n) = n\xi^{n-1}$, then

$$D(\alpha_1 1 + \alpha_2\xi + \alpha_3\xi^2) = \alpha_2 1 + \alpha_2 2\xi. \tag{5.78}$$

The image of $V : 1, \xi, \xi^2$, $\dim(V) = 3$, under differential transformation D is $R : 1, \xi$, $\dim(R) = 2$. The nullspace of D acting on V is $N : 1$, $\dim(N) = 1$. Hence the rank of the transformation, the dimension of its image space or range, is here 2, while the nullity of the transformation is here 1, and rank$D$+nullity$D$=dim$V = 3$. Notice that the range of transformation D and its nullspace are both in V.

Linear operator T is such that, generally

$$T(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1(Tv_1) + \alpha_2(Tv_2) \tag{5.79}$$

sending every element of $V : v_1, v_2$ to vector space $W : Tv_1, Tv_2$, and if w_1, w_2 is another span of W, then

$$\begin{aligned} Tv_1 &= A_{11}w_1 + A_{21}w_2 \\ Tv_2 &= A_{12}w_1 + A_{22}w_2. \end{aligned} \tag{5.80}$$

A linear operator affects a linear transformation between the two vector spaces.

It is by concrete examples that we create generalities, not conversely. Look at the following example of matrix vector spaces. Matrix transposition is a linear operation. If $T(A) = A^T$, then $T(\alpha_1 A_1 + \alpha_2 A_2) = \alpha_1 T(A_1) + \alpha_2 T(A_2)$. For any $A = A(2 \times 2)$

$$A = \alpha_1 \begin{bmatrix} 1 & \\ & \end{bmatrix} + \alpha_2 \begin{bmatrix} & \\ & 1 \end{bmatrix} + \alpha_3 \begin{bmatrix} & 1 \\ & \end{bmatrix} + \alpha_4 \begin{bmatrix} & \\ 1 & \end{bmatrix} = \alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_3 + \alpha_4 A_4 \tag{5.81}$$

where A_1, A_2, A_3, A_4, is the standard basis for $R^{2\times 2}$. Hence $T(A) = \alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_4 + \alpha_4 A_3$, or $T(A) = \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_3 + \beta_4 A_4$, and $\beta_1 = \alpha_1, \beta_2 = \alpha_2, \beta_3 = \alpha_4, \beta_4 = \alpha_3$.

The vector space analogy of functions continuous on the interval $0 \le \xi \le 1$ can be carried further and extended into *normed* spaces by defining for a pair of functions a *scalar* product analogous to $a^T b$, that is related to the geometric notions of length and angle. If ϕ_1 and ϕ_2 are two such functions, then by virtue of the integral linearity

$$\int_0^1 (\alpha_1\phi_1 + \alpha_2\phi_2) d\xi = \alpha_1 \int_0^1 \phi_1 d\xi + \alpha_2 \int_0^1 \phi_2 d\xi. \tag{5.82}$$

Expression

$$(\phi_1, \phi_2) = \int_0^1 \phi_1 \phi_2 d\xi \tag{5.83}$$

behaves exactly like the vector scalar product, and

$$\|\phi\|_\infty = \max_{0\le\xi\le1} |\phi(\xi)| \text{ and } \|\phi\|_2 = \left(\int_0^1 \phi^2 d\xi\right)^{1/2} \tag{5.84}$$

satisfy the conditions on norms or magnitudes.

Both the Cauchy-Schwarz and the triangular inequalities hold for these analogies as they hold for vectors, and in this sense, functions $\phi_1(\xi) = 1$ and $\phi_2(\xi) = \frac{1}{2} - \xi$ are *orthogonal*.

We shall not stray here into more profound abstractions of vector spaces and linear transformations. Suffice it to say that with experience we pass from the concrete to the abstract, and by restricting ourselves to vectors in R^n we concretely created all that is needed to muster the language of vector spaces and transport its concepts into all sorts of analogies and isomorphisms.

exercises

5.7.1. Write a basis for the nullspace of

$$A = \begin{bmatrix} 0 & 1 & 1 \\ & 0 & 1 \\ & & 0 \end{bmatrix}.$$

5.7.2. Find the range and the nullspace of

$$A = \begin{bmatrix} 1 & -1 & 1 & -3 \\ -1 & 1 & -1 & 3 \\ 1 & 1 & 3 & -3 \\ 1 & 2 & 4 & -3 \end{bmatrix}$$

then their intersection.

5.7.3. Let

$$A = A(4 \times 4) = \begin{bmatrix} & 1 & & \\ & & 1 & \\ & & & 1 \\ & & & \end{bmatrix}$$

. Find the nullspace and range of A,A^2,A^3, and their respective intersections.

5.7.4. Prove that the nullspace of AB contains the nullspace of B, and that the column space of AB is in the column space of A.

5.7.5. Explain that $N(AB)$, the nullspace of AB, need not include $N(A)$.

5.7.6. Show that $AB = O$ implies that the columns of B are in the nullspace of A.

5.7.7. Prove that

$$rank(AB) \leq rank(A),\ rank(AB) \leq rank(B)$$

$$rank(A + B) \leq rank(A) + rank(B)$$

and also that

$$nullity\ (AB) \geq nullity\ (B).$$

5.7.8. Prove that if $A = A(m \times n)$ is of rank r, then every submatrix S of A is such that rank $(S) \leq r$, with equality holding for at least one S.

5.7.9. Show that

$$rank(A) = rank(B^{-1}AB).$$

5.7.10. Verify the Frobenius rank inequality

$$rank(AB) + rank(BC) \leq rank(B) + rank(ABC)$$

on

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},\ B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix},\ C = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

5.7.11. Prove that if $N = N(n \times n)$ is a strictly upper-triangular nilpotent matrix,

$$N = \begin{bmatrix} & \times & \times & \times \\ & & \times & \times \\ & & & \times \\ & & & \end{bmatrix}$$

Then nullity(N^2) >nullity(N). Hint: Consider the eventualities of $N_{34} = 1$ and $N_{34} = 0$.

5.7.12. Show that rank$(A^T A) = rank(A)$.

5.8 Orthogonal bases

Vectors $v_1, v_2, \ldots, v_m$ of R^n are mutually orthogonal, or just orthogonal, if pairwise $v_i^T v_j = 0$.

Any n linearly independent vectors in a n-dimensional vector space serve as a basis for that space. No one basis of a vector space has ascendency over any other basis, but orthogonal bases are computationally most desirable.

Theorem 5.27. *An orthogonal set of nonzero vectors, $v_1, v_2, \ldots, v_m$ in R^n is linearly independent. If v is in the space spanned by the m vectors so that $v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m$, then $\alpha_i = v^T v_i / v_i^T v_i$, $i = 1, 2, \ldots, m$.*

Proof. To show linear independence we write $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m = o$, and have upon premultiplication by v_i^T that $v_i^T o = \alpha_i (v_i^T v_i)$. Since $v_i^T v_i \neq 0$ it results that $\alpha_i = 0$.

The m orthogonal vectors are a basis for vector space V, and if $v \in V$, then $v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m$, and $v^T v_i = \alpha_i (v_i^T v_i)$. End of proof.

The theorem demonstrates how easy it is to solve a system of equations with a matrix having orthogonal columns.

If v is not in the subspace spanned by $v_1, v_2, \ldots, v_m$ then $\alpha_i = v^T v_i / v_i^T v_i$ are the coefficients of an orthogonal projection of v into the subspace spanned by the m v's. For if $v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m + v'$, where v' is orthogonal to $v_1, v_2, \ldots, v_m$, then $v_i^T v = \alpha_i v_i^T v_i + v_i^T v' = \alpha_i v_i^T v_i$. On the other hand, if $v_i \in R^m$ $i = 1, 2, \ldots, m$, then $v' = o$, as there are no more than m nonzero orthogonal vectors in R^m.

5.9 Orthogonalization and the QR factorization

Now that we understand how computationally desirable orthogonal bases are we shall learn how to produce them.

The *Gram-Schmidt* orthogonalization algorithm replaces the basis vectors of V one at a time, requiring that each new vector be orthogonal to all those already in place. Let v_1, v_2, v_3, v_4 be the linearly independent span of V. We seek to select in V four nonzero orthogonal vectors q_1, q_2, q_3, q_4 to serve as a new basis for the space, and pick them one after the other in the following sequence:

$$\begin{array}{ll} v_1,\ v_2,\ v_3,\ v_4 & \\ q_1,\ v_2,\ v_3,\ v_4 & q_1 = G_{11} v_1 \\ q_1,\ q_2,\ v_3,\ v_4 & q_2 = G_{12} q_1 + G_{22}\ v_2 \\ q_1,\ q_2,\ q_3,\ v_4 & q_3 = G_{13} q_1 + G_{23} q_2 + G_{33} v_3 \\ q_1,\ q_2,\ q_3,\ q_4 & q_4 = G_{14} q_1 + G_{24} q_2 + G_{34} q_3 + G_{44} v_4. \end{array} \tag{5.85}$$

Vector v_1 is first replaced by $q_1 = G_{11} v_1$, and scalar G_{11} may be selected at will to normalize q_1 so that ${q_1}^T q_1 = 1$. Vector q_2 is sought next in the two dimensional subspace of V spanned by q_1 and v_2, and if $q_2 \neq o$, then by Theorem 5.5 it is linearly independent of v_3 and v_4 that are outside the subspace. The condition that q_2 is orthogonal to q_1, ${q_1}^T q_2 = 0$, produced the scalar equation

$$0 = G_{12} {q_1}^T q_1 + G_{22} {q_1}^T v_2 \quad , \quad G_{12} = -G_{22} \frac{{q_1}^T v_2}{{q_1}^T q_1} \tag{5.86}$$

and G_{22} may be arbitrarily set to render ${q_2}^T q_2 = 1$. Vectors q_1, q_2, v_3, v_4 are linearly independent and still span V. Vector q_3 is now sought in the three-dimensional subspace of V spanned by q_1, q_2, v_3, and any $q_3 \neq o$ is linearly independent of v_4. The conditions that q_3 be orthogonal to q_1 and q_2, namely the conditions that ${q_1}^T q_3 = 0,\ {q_2}^T q_3 = 0$ are stated in the two equations, each with a single unknown

$$\begin{array}{ll} G_{13} {q_1}^T q_1 + G_{33} {q_1}^T v_3 = 0, & G_{23} {q_2}^T q_2 + G_{33} {q_2}^T v_3 = 0 \\ G_{13} = -G_{33} \dfrac{{q_1}^T v_3}{{q_1}^T q_1} & G_{23} = -G_{33} \dfrac{{q_2}^T v_3}{{q_2}^T q_2}, \end{array} \tag{5.87}$$

and G_{33} may be set at will to render ${q_3}^T q_3 = 1$. Lastly q_4 is sought in the whole four-dimensional space and the conditions that ${q_1}^T q_4 = 0,\ {q_2}^T q_4 = 0,\ {q_3}^T q_4 = 0$ reduce, since already ${q_1}^T q_2 = 0,\ {q_1}^T q_3 = 0,\ {q_2}^T q_3 = 0$, to the three linear equations each in only one unknown

$$\begin{array}{lll} G_{14} {q_1}^T q_1 + G_{44} {q_1}^T v_4 = 0, & G_{24} {q_2}^T q_2 + G_{44} {q_2}^T v_4 = 0, & G_{34} {q_3}^T q_3 + G_{44} {q_3}^T v_4 = 0 \\ G_{14} = -G_{44} \dfrac{{q_1}^T v_4}{{q_1}^T q_1} & G_{24} = -G_{44} \dfrac{{q_2}^T v_4}{{q_2}^T q_2} & G_{34} = -G_{44} \dfrac{{q_3}^T v_4}{{q_3}^T q_3} \end{array} \tag{5.88}$$

and we may choose G_{44} so that ${q_4}^T q_4 = 1$.

It is all quite simple, and results in the theoretically important

Theorem 5.28. *Every vector space has an orthonormal basis.*

Proof. Apply the Gram-Schmidt algorithm to any basis of the space. End of proof.

Examples.

$$1.\ \ V\colon \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},\ V\colon \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \alpha_1 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, 3\alpha_1 + \alpha_2 = 0 \tag{5.89}$$

$$2.\ \ V\colon \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad , V\colon \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \tag{5.90}$$

3. Functions $\phi_1(\xi) = 1$, and $\phi_2(\xi) = \xi,\ 0 \le \xi \le 1$ are linearly independent in the sense that $\alpha_1\phi_1 + \alpha_2\phi_2 = 0$ only if $\alpha_1 = \alpha_2 = 0$. Functions $\psi_1(\xi) = \phi_1,\ \psi_2(\xi) = \alpha_1\phi_1 + \phi_2$ are mutually orthogonal in the sense of eq.(5.83) if $2\alpha_1 + 1 = 0$. Then, for any choice of α_1, α_2 there is a unique choice of β_1, β_2, and vice versa, so that $\alpha_1\phi_1 + \alpha_2\phi_2 = \beta_1\psi_1 + \beta_2\psi_2$.

The Gram-Schmidt algorithm also has a matrix formulation.

Let $A = A(m \times n)$ be a matrix with n linearly independent columns. Constructing an orthonormal basis for the column space of A amounts to finding a nonsingular matrix $X(n \times n)$ so that $AX = Q$, where $Q^TQ = I$. Matters are considerably simplified if X can be chosen to be an upper-triangular, or right, matrix.

Theorem 5.29. *Matrix $A(m \times n)$ with linearly independent columns can be uniquely factored as $A = QR$, where $Q(m \times n)$ has orthonormal columns, $Q^TQ = I$, and where $R(n \times n)$ is an upper-triangular matrix with a positive diagonal.*

Proof. Since the columns of A are linearly independent A^TA is positive definite and can be uniquely factored as $A^TA = R^TR, R_{ii} > 0$. Then $Q = AR^{-1}$ is with orthonormal columns since $Q^TQ = R^{-T}A^TAR^{-1} = I$. End of proof.

A *square* matrix with orthonormal columns is said to be *orthogonal.* Then $Q^T = Q^{-1}$, and $Q^TQ = QQ^T = I$. Not only are the columns of square Q orthonormal but so are the rows.

The arithmetics of the Gram-Schmidt orthogonalization of n vectors each with m components takes about mn^2 operations, or n^3 operations if $m = n$. But, if all we want is an orthogonal basis, then the work when $m = n$ is nill since no arithmetic is needed to write an orthogonal basis for R^n.

The Gram-Schmidt orthogonalization algorithm is simple and works fine in integer arithmetic, but it can be erroneous and produce a flawed Q in floating -point computations.

We have already remarked in Sec. 2.12 that if the columns (rows) of A are nearly linearly independent, then the columns (rows) of A^TA are more so. Take, for instance, the matrices

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1+\epsilon \end{bmatrix} \text{ and } A^TA = \begin{bmatrix} 2 & 2+\epsilon \\ 2+\epsilon & 1+(1+\epsilon)^2 \end{bmatrix} \quad |\epsilon| << 1. \tag{5.91}$$

If ϕ and ϕ' denote the angle between the two columns of A and A^TA, respectively, then $\phi = \frac{1}{2}\epsilon$ and $\phi' = \frac{1}{8}\epsilon^2$.

The $A^TA = R^TR$ factorization is in this case with a nearly singular R, and as a result we run the risk that R^{-1} is greatly in error. Matrix $Q = AR^{-1}$, which theoretically has orthonormal columns, is practically far from being orthogonal. In situations where the accurate orthogonality of the columns of Q is crucial, the Gram-Schmidt algorithm is unsatisfactory.

The QR factorization of matrix A is one of the very basic algorithms of linear algebra, and we shall describe next two algorithms, one named after Givens, and the other after Householder, that produce good Q matrices.

Both algorithms use simple orthogonal matrices Q designed so that for given vector x, Qx turns up zeros at specified entries. An orthogonal matrix, we recall, is square and such that $Q^TQ = QQ^T = I$. We also recall that the product of orthogonal matrices is an orthogonal matrix.

The idea of Givens is to use the Jacobi orthogonal rotation matrix

$$J_{kl}=\begin{array}{c} \begin{matrix} & k & & l & \end{matrix} \\ \begin{bmatrix} 1 & & & & \\ & c & & s & \\ & & 1 & & \\ & -s & & c & \\ & & & & 1 \end{bmatrix} \end{array} \ , \ J_{kl}x = \begin{bmatrix} x_1 \\ x_k c + x_l s \\ x_i \\ -x_k s + x_l c \\ x_n \end{bmatrix} \ , \ c^2 + s^2 = 1 \tag{5.92}$$

to enforce $(J_{kl}x)_l = -x_k s + x_l c = 0$ with

$$c = x_k/(x_k^2 + x_l^2)^{1/2} \ , \ s = x_l/(x_k^2 + x_l^2)^{1/2}. \tag{5.93}$$

One multiplication of x by J sets one entry of Jx equal to zero. Successive multiplications by the orthogonal matrices create more zeroes. Because $\|Jx\| = \|x\|$, $\|x\| = \sqrt{x^T x}$, at least one entry in Jx must remain nonzero.

Let $A = A(m \times n),\ m \geq n$, be the matrix we want to factor. Premultiplication of A by J_{kl} affects the k and l rows of A only, and we adjust c and s to create a permanent zero at $(J_{kl}A)_{lk}$. Repeated premultiplication by the Jacobi rotation matrices renders the product upper-triangular (trapezoid), as described below:

$$\begin{matrix} A_0 = A & & A_1 = J_{12}A_0 & & A_2 = J_{13}A_1 & & A_3 = J_{14}A_2 \\ \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ \times & \times & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \end{matrix} \tag{5.94}$$

$$\begin{matrix} & A_4 = J_{23}A_3 & & A_5 = J_{24}A_4 & & A_6 = J_{34}A_5 \\ \rightarrow & \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & \times & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & 0 \end{bmatrix}. \end{matrix}$$

Since the product of orthogonal matrices is an orthogonal matrix we write $Q^T = J_{34}J_{24}$ $J_{23}J_{14}J_{13}J_{12}$ and have that $Q^T A = R$, or $A = QR$. In our example the last row of A_6 is zero and therefore

$$
\begin{array}{ccccccc}
A & & Q & R & & & \\
\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} & = & \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix} & \begin{bmatrix} \times & \times & \times \\ & \times & \times \\ & & \times \\ & & \end{bmatrix} & = & \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} & \begin{bmatrix} \times & \times & \times \\ & \times & \times \\ & & \times \end{bmatrix}.
\end{array} \tag{5.95}
$$

Givens' method can factor A even with linearly dependent columns, except that then R is of type 0. For example:

$$
\frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \sqrt{2} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \tag{5.96}
$$

and it therefore produces Q matrices that have nearly perfect orthogonal columns. Whenever the orthogonality of Q is important, the method of Givens is preferred to that of Gram-Schmidt. The fact that Q is nearly orthogonal does not mean, however, that it is near the theoretical factor, nor that R is.

In the same way a finite sequence of premultiplications of A by the elementary orthogonal Jacobi matrices leads to the factorization $Q^T A = L$ with a lower-triangular L, as described below:

$$
\begin{array}{ccccccc}
A & & J_{23}A & & J_{13}J_{23}A & & J_{12}J_{13}J_{23}A \\
\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & \times & \times \\ \times & \times & 0 \\ \times & \times & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & \times & 0 \\ \times & \times & 0 \\ \times & \times & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & 0 & 0 \\ \times & \times & 0 \\ \times & \times & \times \end{bmatrix}.
\end{array} \tag{5.97}
$$

To create one zero in Qx by the method of Givens requires four arithmetical operations and one square root, whereby the complete factorization of $A = A(m \times n)$ requires about $2n^2(m - n/3)$ operations plus $n(2m - n)/2$ square roots, but it produces Q in a factored form.

For a geometrical interpretation of Givens' method imagine $x = [x_1 \ x_2 \ x_3]^T$ being first rotated in space by J_{12} into $x' = J_{12}x = [x_1' \ 0 \ x_3']^T$ in the x_1x_3 plane, then by J_{13} into $x'' = J_{13}x' = [x_1'' \ 0 \ 0]^T = \sigma\|x\|[1 \ 0 \ 0]^T$, where $\sigma = \pm 1$, along the x_1 axis.

Computational efficiency is increased if the successive rotations to turn x colinear with e_1 are combined into a single rotation. Geometrically it is obvious what this rotation could

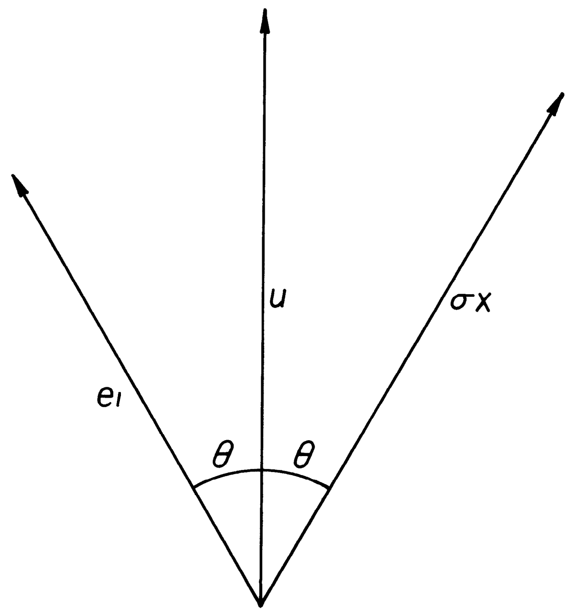

Fig. 5.1

be. Figure 5.1 shows e_1 and σx, with $\sigma = |x_1|/x_1$, so as to create an acute angle between e_1 and σx. Rotation of σx by 180^o around axis u in the plane of σx and e_1, that bisects angle $2\theta, \cos 2\theta = |x_1|/\|x\|$, between them brings σx into colinearity with e_1.

This is also true in R^n, and for a 180^o rotation we use matrix

$$R = -I + 2uu^T \tag{5.98}$$

that is such that $Ru = u,\ Rv = -v,\ v^Tu = 0$, with

$$u = \beta(e_1 + \frac{\sigma}{\|x\|}x),\ \sigma = |x_1|/x_1 \tag{5.99}$$

where β is to make $u^Tu = 1$, or

$$1 = 2\beta^2(1 + |x_1|/\|x\|). \tag{5.100}$$

We readily verify that $Rx = \sigma\|x\|e_1$.

It is customary to use the reflection matrix H (after Householder) $H = -R = I - 2uu^T$ instead of rotation matrix R .

As an exercise the reader could undertake writing the rotation matrix R for an axis that is orthogonal to both e_1 and x.

What H did to the entire vector x it can do to a portion of it by partitioning . Say we want to leave the first k entries of x untouched but want to have $(Hx)_i = 0$ for $i = k+2$, $k+3, \ldots, n$. To achieve this we write $u = [0\ 0\ \ldots\ 0, u_{k+1}\ \ldots\ u_n]^T = [o^T\ v^T]^T, v^T v = 1$, so that H assumes the partitioned form

$$H = I - 2uu^T = \begin{bmatrix} I & O \\ O & I - 2vv^T \end{bmatrix} \tag{5.101}$$

and if $x = [x_1^T\ x_2^T]^T$, then

$$Hx = \begin{bmatrix} x_1 \\ (I - 2vv^T)x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ H'x_2 \end{bmatrix}. \tag{5.102}$$

It is now obvious how to apply the Householder reflection matrices to create the QR factorization of A:

$$\begin{array}{ccccccc} A_0 = A & & A_1 = H_1 A_0 & & A_2 = H_2 A_1 & & A_3 = H_3 A_2 \\ \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{bmatrix} & \rightarrow & \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & 0 \end{bmatrix} \end{array} \tag{5.103}$$

where

$$H_2 = \begin{bmatrix} 1 & \\ & H_2' \end{bmatrix}, \quad H_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & H_3' \end{bmatrix}. \tag{5.104}$$

As with Givens' method, Householders' method can also factor A, even with linearly dependent columns, and produces therefore a nearly perfect orthogonal matrix Q. Round-off errors are not manifested so much in loss of orthogonality as in the computed Q being rotated relative to the theoretical one.

Householders' method is faster than Givens' method; the number of operations in this method are approximately $n^2(m - n/3)$ plus $\frac{1}{2}n$ square roots.

exercises

5.9.1. Use the Gram -Schmidt algorithm to write an orthogonal basis for $V : v_1 = [1\ 1\ 1]^T$, $v_2 = [1\ \ -1\ 1]^T$.

5.9.2. Use the Gram-Schmidt orthogonalization to write an orthogonal basis for $V: \ [1\ 1\ 1\ 1]^T, \ [1\ 1\ 1\ 0]^T, \ [1\ 1\ 0\ 0]^T$.

5.9.3. Let V be a subspace of R^4 spanned by $v_1 = [1\ 1\ 1\ 1]^T, \ v_2 = [1\ 1\ 1\ -1]^T$. Write a basis for the orthogonal complement of V.

5.9.4. Write an orthogonal basis for R^4 starting with $v_1 = [1\ 1\ 1\ 1]^T$.

5.9.5. Write the QR factorization of

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

5.9.6. For matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.095 \end{bmatrix}$$

find upper-triangular R so that $A^TA = R^TR$, then form $Q = AR^{-1}$. Use a four-digit round-off floating-point arithmetic. Show that the computed $Q = [q_1\ q_2]$ is not orthogonal; $\sqrt{q_2^Tq_2} = 1.223$ with the angle of 91^o between q_1 and q_2.

5.9.7. Use a succession of Jacobi rotation matrices to produce orthogonal matrix Q so that $Qx = \alpha e_1$ for $x = [1\ 1\ 1\ 1]^T$. Do the same with the Householder H matrix.

5.9.10. Write an orthogonal basis for $F: \ 1, \xi, \xi^2, \ |\xi| \leq 1$.

5.9.11. What is R in the QR factorization of A such that $A^TA = D$ is diagonal?

5.9.12. Show that the orthogonal complement of $V \cup W$ is the intersection of the orthogonal complements of V and W. Think first geometrically.

5.9.13. Verify that the provisions of Theorem 5.32 hold true for

$$A = \begin{bmatrix} 1 & 3 & 2 & 1 & 3 \\ -1 & 2 & 3 & 4 & 7 \\ 1 & -1 & -2 & -3 & -5 \\ -1 & 0 & 1 & 2 & 3 \end{bmatrix}.$$

5.10 Orthogonal projections

In this section we take up the question of vector approximations.

Definition. *Let W be a subspace of vector space V. A best approximation to a given $v \in V$ from among the vectors of W is a nonzero element $w' \in W$ such that*

$$\|v - w'\| \leq \|v - w\| \tag{5.105}$$

for every $w \in W$.

We restrict ourselves to the ℓ_2 norm, $\|v\| = (v^T v)^{1/2}$.

Theorem 5.30. *Let W be a subspace of V and $v \in V$ be given. Then:*

1. A best approximation w' to v exists if and only if there is at least one $w \in W$ for which $v^T w \neq 0$.

2. A best approximation, when it exists, is unique.

3. Vector $w' \in W$ is a best approximation to v if and only if $w^T(v - w') = 0$ for every $w \in W$.

4. If $w_1, w_2, \ldots, w_m$ is an orthonormal basis for W, then the best approximation to v is

$$w' = (w_1^T v)w_1 + (w_2^T v)w_2 + \cdots + (w_m^T v)w_m. \tag{5.106}$$

Proof. Let $v_1, v_2, \ldots, v_n$ be an orthonormal basis for V, and $v_1, v_2, \ldots, v_m$, $m < n$ an orthonormal basis for W. If

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m + \cdots + \alpha_n v_n \tag{5.107}$$

and

$$w = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_m v_m \tag{5.108}$$

then

$$v - w = (\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \cdots + (\alpha_m - \beta_m)v_m \tag{5.109}$$

$$+\alpha_{m+1}v_{m+1} + \cdots + \alpha_n v_n$$

and

$$\|v-w\|^2 = (\alpha_1-\beta_1)^2 + (\alpha_2-\beta_2)^2 + \cdots + (\alpha_m-\beta_m)^2 \\ + \alpha_{m+1}^2 + \cdots + \alpha_n^2. \tag{5.110}$$

Each term on the right is nonnegative, and $\|v-w\|^2$ possesses a unique minimum with respect to $\beta_1, \beta_2, \ldots, \beta_m$ at

$$\alpha_1 - \beta_1 = 0,\ \alpha_2 - \beta_2 = 0, \ldots, \alpha_m - \beta_m = 0 \tag{5.111}$$

and the best approximation in W is

$$w' = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m \tag{5.112}$$

so that

$$\|v-w'\|^2 = \alpha_{m+1}^2 + \cdots + \alpha_n^2. \tag{5.113}$$

A nonzero best approximation exists if and only if at least one of $\alpha_1, \alpha_2, \ldots, \alpha_m$ is nonzero, or v is not orthogonal to W; we have therefore proved statements 1 and 2.

To prove statements 3 and 4, notice that $v - w' = \alpha_{m+1}v_{m+1} + \cdots + \alpha_n v_n$ is orthogonal to $v_1, v_2, \ldots, v_m$, and hence to every $w \in W$. This condition is necessary and sufficient for w' to be the best approximation. The condition $w_i^T(v-w') = 0\ i = 1, 2, \ldots, m$ yields 4. End of proof.

Vector w', the best approximation to v in W, is the *orthogonal projection* of v into W.

Orthogonal bases are most convenient here but they are not always directly available. We shall consider now the question of best approximation under less favorable circumstances.

Theorem 5.31. *Let $\phi_1(x) = x^Ta$, and $\phi_2(x) = x^TAx$ be a linear and a quadratic form for $x \in R^n$, respectively. Then*

$$\frac{\partial \phi_1}{\partial x} = a,\ \frac{\partial \phi_2}{\partial x} = Ax + A^Tx \tag{5.114}$$

and if $A = A^T$, then

$$\frac{\partial \phi_2}{\partial x} = 2Ax \tag{5.115}$$

where

$$\frac{\partial \phi}{\partial x} = \left[\frac{\partial \phi}{\partial x_1}\ \ \frac{\partial \phi}{\partial x_2} \cdots \frac{\partial \phi}{\partial x_n}\right]^T \tag{5.116}$$

is the gradient of $\phi = \phi(x)$.

Proof. Write

$$\phi_1(x) = \sum_{i=1}^{n} x_i a_i \tag{5.117}$$

and obtain $\partial\phi_1/\partial x_i = a_i$. Then write

$$\phi_2(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j \tag{5.118}$$

and obtain

$$\frac{\partial \phi_2}{\partial x_i} = \sum_{j=1}^{n} (A_{ij} + A_{ji}) x_j. \tag{5.119}$$

End of proof.

Let $v \in R^m$ be a given vector and consider the search for $w \in W$ so that

$$\phi(w) = (v - w)^T (v - w) = \|v - w\|^2 \tag{5.120}$$

is minimal. Let $a_1, a_2, \ldots, a_n$, the columns of $A = A(m \times n)$, be a basis for W so that we may write $w = Ax$ $x \in R^n$, $n \leq m$. Objective function $\phi(w)$ becomes with this

$$\phi(x) = (v - Ax)^T (v - Ax) = v^T v - 2x^T A^T v + x^T A^T Ax. \tag{5.121}$$

A necessary condition for minimum ϕ is that the gradient of $\phi(x)$ vanish, or

$$\frac{\partial \phi}{\partial x} = -2A^T v + 2A^T Ax = o \tag{5.122}$$

and $x = (A^T A)^{-1} A^T v$. An inverse to $A^T A$ exists since the columns of A are linearly independent, and if $A^T v \neq o$, then a nonzero x for which $\partial\phi/\partial x = o$ exists. Matrix $(A^T A)^{-1} A^T$ is occasionally called the *generalized inverse* of A.

To prove that $s = (A^T A)^{-1} A^T v$ *minimizes* $\phi(x)$ we rewrite it as

$$\phi(x) = (x - s)^T A^T A(x - s) + v^T (v - As) \tag{5.123}$$

and conclude that, since $(x - s)^T A^T A(x - s) \geq 0$ with equality holding only when $x = s$, s is the unique minimizer of $\phi(x)$.

Example. To find the best approximation to $\phi(\xi) = e^{\xi}$ among $\psi(\xi) = \alpha_1 + \alpha_2\xi \;\; 0 \leq \xi \leq 1$ in the sense that

$$\|\phi - \psi\|^2 = \int_0^1 (\alpha_1 + \alpha_2\xi - e^{\xi})^2 d\xi \tag{5.124}$$

is minimal.

Setting the gradient of $\|\phi - \psi\|^2$ equal to zero we obtain

$$\int_0^1 (\alpha_1 + \alpha_2\xi - e^{\xi})1 d\xi = 0 \;,\; \int_0^1 (\alpha_1 + \alpha_2\xi - e^{\xi})\xi d\xi = 0 \tag{5.125}$$

or after integration

$$\alpha_1 + \frac{1}{2}\alpha_2 = e - 1 \;,\; \frac{1}{2}\alpha_1 + \frac{1}{3}\alpha_2 = 1 \tag{5.126}$$

and $\alpha_1 = 4e - 10,\; \alpha_2 = 18 - 6e$. Best approximating among $\psi(\xi)$ is

$$\psi_{min}(\xi) = 0.8731273 + 1.690309\xi \tag{5.127}$$

for which $\|\phi - \psi_{min}\| = 0.0628$.

The choice of norm, or the way the question of best approximation is posed, has a significant influence on the ease with which ψ_{min} is computed. Computation of the best approximation in the L_2 norm of eq. (5.124) is, at least in principle, simple. Finding the best approximation in the L_∞, or uniform, norm is an altogether different matter. Here we are confronted with the min-max problem of finding

$$\min_{\alpha_1,\alpha_2} \max_{0\leq\xi\leq 1} |\phi - \psi| \tag{5.128}$$

and this is far from simple. We cannot say more here on this difficult subject, which belongs to the theory of polynomial approximations, except that lengthy computation yields

$$\psi_{min}(\xi) = 0.894067 + 1.718282\xi \tag{5.129}$$

and

$$\max_{0\leq\xi\leq 1} |\psi_{min} - e^{\xi}| = 0.105933. \tag{5.130}$$

exercises

5.10.1. Show that $A(A^TA)^{-1}A^T$, $A = A(m \times n)$, $m > n$, is an orthogonal projection matrix.

5.10.2. Show that if P is an orthogonal projection matrix ($P = P^T, P^2 = P, P \neq I$), then $I - P$ is also a projection matrix, and that $Q = I - 2P$ is orthogonal.

5.11 Orthogonal subspaces

The row and column spaces of a rectangular matrix have interesting orthogonality relationships.

Definition. *Two vector subspaces V and W are orthogonal if $v^Tw = 0$ holds for every $v \in V$ and every $w \in W$. If V and W are both subspaces of R^n, and $\dim(V) + \dim(W) = n$, then V and W are orthogonal complements of R^n.*

Obviously orthogonal subspaces are disjoint. Only the zero vector is orthogonal to itself.

Theorem 5.32.

1. Let $C(A)$ and $N(A^T)$ be the column space or range of $A = A(m \times n)$, and the nullspace of $A^T = A^T(n \times m)$, respectively. They are orthogonal complements of R^m.

2. Let $R(A)$ and $N(A)$ be the row space and the nullspace of $A = A(m \times n)$, respectively. They are orthogonal complements of R^n.

Proof.

1. First we show that every $x \in N(A^T)$ is orthogonal to every $x'' \in C(A)$. Indeed, if $A^Tx = o$, and $Ax' = x''$, $x' \in R^n$, then $x^Tx'' = x^TAx' = x'^TA^Tx = 0$. Secondly, we have that

$$\dim C(A^T) + \dim N(A^T) = m \tag{5.131}$$

and since $\dim C(A^T) = \dim C(A)$ it results that

$$\dim C(A) + \dim N(A^T) = m. \tag{5.132}$$

2. As for 1, except that here

$$\dim C(A) + \dim N(A) = n \tag{5.133}$$

and $\dim C(A) = \dim C(A^T) = \dim R(A)$. End of proof.

Corollary 5.33. *If $A(n \times n) = A^T(n \times n)$, then $N(A)$ and $C(A)$ are orthogonal complements of R^n.*

Proof. Set $A = A^T$ in Theorem 5.32. End of proof.

Example. The four fundamental spaces of

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & 4 \\ -1 & -2 & -3 \\ 3 & 4 & 5 \end{bmatrix} \tag{5.134}$$

are

$$\text{column space } C(A) : \begin{bmatrix} 3 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -2 \\ 4 \end{bmatrix}, \quad \text{row space } R(A) : \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix},$$

$$\text{nullspace } N(A^T) : \begin{bmatrix} -1 \\ 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ 0 \\ 3 \end{bmatrix}, \quad \text{and nullspace } N(A) : \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}; \tag{5.135}$$

and

$$\dim C(A) = \dim R(A),\ \dim R(A) + \dim N(A^T) = 4,\ \dim R(A) + \dim N(A) = 3 \tag{5.136}$$

and

$$C(A) \perp N(A^T) \ \ , \ \ R(A) \perp N(A) \tag{5.137}$$

where $\perp$ means orthogonal to.

exercises

5.11.1. Write the 2×2 matrices that have an equal column and nullspace. Do the same for 3×3 matrices.

5.11.2. Prove that if $A = A(n \times n)$ is such that its column space equals its nullspace, $R(A) = N(A)$, as in

$$A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$$

then $A + A^T$ is nonsingular. Can this happen for a 3×3 matrix?

5.11.3. Consider the set of vectors a_1, a_2, a_3, a_4 in R^n. Define $V_1 = |a_1|, V_2 = V_1|a_2|sin\theta_2$, where θ_2 is the angle between a_2 and its orthogonal projection upon a_1. In this way V_1 is the length of a_1 and V_2 is the area of the parallelogram formed by a_1 and a_2. Continue in this manner and write $V_3 = V_2|a_3|sin\theta_3 = |a_1||a_2||a_3|sin\theta_2 sin\theta_3$, where θ_3 is the angle between a_3 and its orthogonal projection upon the space spanned by a_1, a_2. If a_1, a_2, a_3 are in R^3, then V_3 is the volume of the parallelepiped formed by a_1, a_2, a_3, and we know that $V_3 = |det[a_1 \ a_2 \ a_3]|$. What is V_4 defined recursively by $V_4 = V_3|a_4|sin\theta_4$, with a_1, a_2, a_3, a_4 in R^4 ? What is generally $V_n = V_{n-1}|a_n|sin\theta_n$ with $a_1, a_2, \ldots, a_n$ in R^n ?

Answers

section 5.1 5.1.1. No.

5.1.2. Yes. Yes.

5.1.3. Yes.

5.1.4. No.

5.1.6. $v_1 = -2w_2 + w_3, v_2 = w_2 + w_3; w_1 = v_1 + v_2, w_2 = -\frac{1}{3}v_1 + \frac{1}{3}v_2, w_3 = \frac{1}{3}v_1 + \frac{2}{3}v_2.$

section 5.2

5.2.1. No: $v_1 - v_2 + v_3 = o.$

5.2.2. Yes.

5.2.3. $7\alpha^2 - 18\alpha + 8 = 0.$

5.2.4. $\beta^3 + 1 \neq 0.$

5.2.8 If $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 = o$, then $\alpha_1 = \alpha_3 - \alpha_4 + \alpha_5, \alpha_2 = -\alpha_3 - \alpha_4 - 2\alpha_5$, and we may set $\alpha_4 = \alpha_5 = 0, \alpha_3 \neq 0$ to have $v_1 - v_2 + v_3 = o$.

section 5.3

5.3.1. $x_1 - x_2 + x_3 = 0,\ x = x_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$

5.3.2. $x = [x_1\ x_2\ x_3\ x_4]^T,\ x_1 \neq x_2, x_3 \neq x_4.$

5.3.3. $V \cap W : [1\ \ -1\ 1\ \ -1]^T.$

section 5.6

5.6.2.

$$N(A) : \begin{bmatrix} 1 \\ \\ \\ -3 \\ -1 \end{bmatrix}, \begin{bmatrix} \\ 1 \\ \\ -5 \\ -3 \end{bmatrix}, \begin{bmatrix} \\ \\ 1 \\ 8 \\ 5 \end{bmatrix}, \quad N(A^T) : \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

$$R(A): \begin{bmatrix} 4 \\ \\ 1 \\ 3 \\ -5 \end{bmatrix}, \begin{bmatrix} \\ 4 \\ -7 \\ -1 \\ 3 \end{bmatrix}, \quad C(A): \begin{bmatrix} 1 \\ \\ -1 \end{bmatrix}, \begin{bmatrix} \\ 1 \\ 2 \end{bmatrix}.$$

.

5.6.5. Yes. Yes.

5.6.9. $AX = XA, X = \alpha_1 \begin{bmatrix} 1 & & \\ & & \\ & & 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} & & \\ & 1 & \\ & & \end{bmatrix} + \alpha_3 \begin{bmatrix} & & 1 \\ & & \\ & & \end{bmatrix}.$

5.6.10. No.

5.6.12.

$$Y = 2X_{12} \begin{bmatrix} & 1 \\ & \end{bmatrix} - 2X_{21} \begin{bmatrix} & \\ 1 & \end{bmatrix}. \text{ If } X = X_{11} \begin{bmatrix} 1 & \\ & \end{bmatrix} + X_{22} \begin{bmatrix} & \\ & 1 \end{bmatrix}, \text{ then } Y = O.$$

$$Y = \begin{bmatrix} 3X_{11} & 2X_{12} \\ 4X_{21} & 3X_{22} \end{bmatrix}. \text{ Range: } X = X. \text{ Null space: } X = O.$$

section 5.7

5.7.1. $N(A) : e_1$.

5.7.2. Ans.

$$C(A): \begin{bmatrix} 2 \\ -2 \\ \\ -1 \end{bmatrix}, \begin{bmatrix} \\ \\ 2 \\ 3 \end{bmatrix}; \; N(A): \begin{bmatrix} 3 \\ \\ \\ 1 \end{bmatrix}, \begin{bmatrix} \\ 3 \\ -3 \\ -2 \end{bmatrix}; \; C(A) \cap N(A): \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}.$$

5.7.3. $C(A) : e_1, e_2, e_3;\ N(A) : e_1;\ C(A) \cap N(A) : e_1$. $C(A^2) : e_1, e_2;\ N(A^2) : e_1, e_2;\ C(A^2) = N(A^2)$. $C(A^3) : e_1,\ N(A^3) : e_1, e_2, e_3;\ C(A^3) \cap N(A^3) : e_1$.

section 5.9

5.9.1. $q_1 = v_1, q_2 = v_2 + \alpha v_1, \alpha = -1/3$, $V : [1\ 1\ 1]^T, [1\ -2\ 1]^T$.

5.9.2. $q_1 = [1\ 1\ 1\ 1]^T,\ q_2 = [1\ 1\ 1\ -3]^T,\ q_3 = [1\ 1\ -2\ 0]^T$.

5.9.3. $v_3 = [1\ 0\ -1\ 0]^T,\ v_4 = [0\ 1\ -1\ 0]^T$, $R^4 : v_1, v_2, v_3, v_4$.

5.9.4. $v_1 = [1\ 1\ 1\ 1]^T,\ v_2 = [1\ 1\ \ -1\ \ -1]^T.\ v_3 = [-1\ 1\ 1\ \ -1]^T,\ v_4 = [1\ \ -1\ 1\ \ -1]^T.$

5.9.5.

$$A = \frac{1}{6}\begin{bmatrix} 3 & -3r \\ 3 & r \\ 3 & r \\ 3 & r \end{bmatrix}\begin{bmatrix} 2 & 1 \\ & r \end{bmatrix},\ r = \sqrt{3}.$$