M341 (56140), 11/19/2012

Ravi Srinivasan (rav@math.utexas.edu)

Table of contents

1 Span, linear independence, basis and dimension

This section summarizes the concepts of span, linear independence, basis and dimension as discussed in lecture (Sections 4.3-4.6 in Andrilli and Hecker). We try to keep notation similar to that of the book but will deviate from it slightly. Also note that we will *not* use the simplified span method described in the book.

In all that follows, $\mathcal{V}$ will denote a given vector space. In the special case when $\mathcal{V} = \mathbb{R}^m$, we will be able to use matrix methods to answer several questions regarding span and linear independence of a given set of vectors. If $\mathcal{V}$ is a general vector space, matrix methods can be used after fixing a basis of $\mathcal{V}$ and writing all vectors in coordinates of this basis. This is what is known as coordinatization, which we will cover in Section 2 (Section 4.7 in Andrilli and Hecker).

1.1 Span

To begin, we introduce the notion of a linear combination of a finite set of vectors:

Definition. *A **linear combination** of the vectors* $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subseteq \mathcal{V}$ *is any vector of the form* $a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n$ *for scalars* $a_1, \ldots, a_n \in \mathbb{R}$.

Definition. *Let S be a subset (possibly infinite) of $\mathcal{V}$. If $S = \{\}$ we define* $\operatorname{span}(S) = \{\mathbf{0}\}$. *If S is nonempty we define*

$$\operatorname{span}(S) = \{a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n: \quad \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in S, \quad a_1, \ldots, a_n \in \mathbb{R}\}$$

*That is, the **span** of S is the set of all (finite) linear combinations of elements of S.*

Definition. *The span of the rows of a matrix is called its **row space** while the span of the columns of a matrix is called its **column space**.*

In lecture, we proved the following theorem and corollary (see p. 231-232 in Andrilli & Heckler):

Theorem 1. *Let S be a nonempty subset of $\mathcal{V}$. Then*

1. $S \subseteq \operatorname{span}(S)$
2. $\operatorname{span}(S)$ *is a subspace of* $\mathcal{V}$
3. *If* $\mathcal{W}$ *is a subspace of* $\mathcal{V}$ *with* $S \subseteq \mathcal{W}$ *then* $\operatorname{span}(S) \subseteq \mathcal{W}$
4. $\operatorname{span}(S)$ *is the smallest subset of* $\mathcal{V}$ *containing* S.

Corollary. *Let S_1 and S_2 be subsets of $\mathcal{V}$ with $S_1 \subseteq S_2$. Then* $\operatorname{span}(S_1) \subseteq \operatorname{span}(S_2)$.

1.1.1 Computing span in $\mathbb{R}^m$

Given a finite set of vectors $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subseteq \mathbb{R}^m$, how can we determine which vectors $\boldsymbol{b} \in \mathbb{R}^m$ lie in the span of S? To do this we reduce this to an equivalent problem of determining existence of solutions to a system of equations.

Theorem 2. *Let A be the matrix which has $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subseteq \mathbb{R}^m$ as its columns. Then $\boldsymbol{b} \in \operatorname{span}(S)$ if and only if the system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent (i.e., there exists at least one solution $\boldsymbol{x}$).*

Proof. We know that $\boldsymbol{b} \in \operatorname{span}(S)$ if there exist scalars $a_1, \ldots, a_n$ such that $a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n = \boldsymbol{b}$. Letting $\boldsymbol{x} = [a_1, \ldots, a_n]^T$, this is equivalent to showing existence of solutions to $A\boldsymbol{x} = \boldsymbol{b}$. That is, we can find at least one solution $\boldsymbol{x}$ if and only if $\operatorname{rref}[A|\boldsymbol{b}]$ has no rows of the form $\left[\begin{array}{ccc|c} 0 & \cdots & 0 & r \end{array}\right]$ with $r \neq 0$.

□

Corollary. $\operatorname{span}(S) = \mathbb{R}^m$ *if and only if* $\operatorname{rref}(A)$ *has a pivot in every row. Furthermore, if $n < m$ (i.e., the number of vectors n in S is strictly less than the dimension m of $\mathbb{R}^m$) then S cannot span $\mathbb{R}^m$.*

Proof. $A\boldsymbol{x} = \boldsymbol{b}$ is consistent for every $\boldsymbol{b} \in \mathbb{R}^m$ if and only if there are no zero rows in $\operatorname{rref}(A)$—that is, if there is a pivot in every row of $\operatorname{rref}(A)$. In the case when $n < m$, we must have zero rows in $\operatorname{rref}(A)$ since A has more rows than columns. Using the previous theorem completes the proof.

□

1.1.2 Minimal spanning subset

Often we are given a set of vectors that are redundant, in that there is a smaller set $B \subseteq S$ such that $\operatorname{span}(B) = \operatorname{span}(S)$ (i.e., we can discard some of the vectors in S without changing the span). This leads us to the following definition.

Definition. *Let $S \subseteq \mathcal{V}$. We say that B is a **minimal spanning subset** of S if these two properties hold:*

i. $\operatorname{span}(B) = \operatorname{span}(S)$

ii. *If $C \subset B$ with $C \neq B$, then* $\operatorname{span}(C) \neq \operatorname{span}(S)$.

Given $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subseteq \mathbb{R}^m$, how we compute a minimal spanning subset B? We can use a system of equations to find an answer as before. Consider the equation $A\boldsymbol{x} = \boldsymbol{b}$, where A is the matrix with elements of S as its columns as before and $\boldsymbol{b}$ is an arbitrary vector in $\operatorname{span}(S)$. Since $\boldsymbol{b}$ is in the span of S, we know that $A\boldsymbol{x} = \boldsymbol{b}$ has *at least* one solution $\boldsymbol{x}$, and certainly more than one solution if $\operatorname{rref}(A)$ has columns with no pivot (since this yields free variables and non-uniqueness of solutions to $A\boldsymbol{x} = \boldsymbol{0}$). Note now that we only needed one solution to $A\boldsymbol{x} = \boldsymbol{b}$ in order for $\boldsymbol{b} \in \operatorname{span}(S)$, so we can pick the one in which all the free variables are set equal to 0. However, this is equivalent to writing $\tilde{A}\tilde{\boldsymbol{x}} = \boldsymbol{b}$, where $\tilde{A}$ is the smaller matrix consisting of only the pivot columns of A and $\tilde{\boldsymbol{x}}$ is the smaller vector consisting of only nonzero entries of the solution $\boldsymbol{x}$ (i.e., those that did not correspond to free variables). To summarize, we have proven:

Theorem 3. *Let A be the matrix which has $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subseteq \mathbb{R}^m$ as its columns. Then a minimal spanning subset of S is given by the pivot columns of A.*

Example. Suppose $S = \{[1,1]^T, [-2,-1]^T, [1,0]^T, [0,1]^T\}$. To compute a minimal spanning subset of these vectors, we let $A = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}$ and find that $\operatorname{rref}(A) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 1 \end{bmatrix}$. Since $\operatorname{rref}(A)$ has pivots in its first and second columns, the pivot columns $B = \{[1,1]^T, [-2,-1]^T\}$ of A comprise a minimal spanning subset of S.

Note that minimal spanning subsets are not unique—for example, in the previous example the set $\tilde{B} = \{[1,0]^T, [0,1]^T\}$ also constitutes such a set.

1.2 Linear independence

We begin by defining the notion of linear independence for finite sets of vectors.

Definition. *Let $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be a finite subset of $\mathcal{V}$. Then S is **linearly independent** if and only if the equation $a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n = \mathbf{0}$ implies that $a_1 = a_2 = \cdots = a_n = 0$. If S is not linearly independent, we say S is **linearly dependent**. We will define the empty set $\{\}$ to be linearly independent.*

We will extend this definition to infinite sets as follows:

Definition. *An infinite subset $S \subseteq \mathcal{V}$ is linearly independent if and only if every finite subset T of S is linearly independent.*

It should be fairly obvious that any set that contains the zero vector must be linearly dependent. We will now state some alternate characterizations of linear independence (see p. 243-246 in Andrilli & Heckler for further discussion and proofs). The first equivalence simply expresses the fact that linearly independent vectors cannot "overlap," in the sense that any one vector cannot be written as a linear combination of the others. The second equivalence expresses the same fact in an iterative manner, while the last equivalence demonstrates that linear independence yields a unique way of writing any vector in the span:

Theorem 4. *Let $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be a finite subset of $\mathcal{V}$. Then linear independence of S is equivalent to each of the following statements:*

1. *There is no vector $\boldsymbol{v} \in S$ such that $\boldsymbol{v} \in \operatorname{span}(S - \{\boldsymbol{v}\})$.*
2. *$\boldsymbol{v}_1 \neq \mathbf{0}$ and for each $k = 2, \ldots, n$, $\boldsymbol{v}_k \notin \operatorname{span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1}\})$.*
3. *Every vector $\boldsymbol{w} \in \operatorname{span}(S)$ can be expressed uniquely as a linear combination of elements of S.*

Proof. We only prove the third equivalence. Suppose S is linearly independent and $\boldsymbol{w} \in \operatorname{span}(S)$. If we can write $\boldsymbol{w} = a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n$ and $\boldsymbol{w} = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n$, then subtracting the second equality from the first implies $(a_1 - c_1)\boldsymbol{v}_1 + \cdots + (a_n - c_n)\boldsymbol{v}_n = \mathbf{0}$. Therefore, by linear independence of S we have $a_1 - c_1 = \cdots = a_n - c_n = 0$, which shows that $\boldsymbol{w}$ is given by a unique linear combination of S. For the converse, suppose that every vector $\boldsymbol{w} \in \operatorname{span}(S)$ can be expressed uniquely as a linear combination of S. Since $\mathbf{0} \in \operatorname{span}(S)$, there is only one linear combination $a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n$ of S that equals $\mathbf{0}$. Since $0\boldsymbol{v}_1 + \cdots + 0\boldsymbol{v}_n = \mathbf{0}$ this means that $a_1 = a_2 = \cdots = a_n = 0$ and we have linear independence of S.

□

1.2.1 Determining linear independence in $\mathbb{R}^m$

Given a finite set of vectors $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subseteq \mathbb{R}^m$, how can we determine if S is linearly independent? We do this by reducing to an equivalent problem of proving uniqueness of solutions to a system of equations.

Theorem 5. *Let A be the matrix which has $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subseteq \mathbb{R}^m$ as its columns. Then S is linearly independent if and only if the system $A\boldsymbol{x} = \boldsymbol{0}$ has no nontrivial solutions (i.e., $\boldsymbol{x} = \boldsymbol{0}$ is the unique solution).*

Proof. S is linearly independent if $a_1\boldsymbol{v}_1 + \cdots + a_n\boldsymbol{v}_n = \boldsymbol{0}$ implies $a_1 = a_2 = \cdots = a_n = 0$. Letting $\boldsymbol{x} = [a_1, \ldots, a_n]^T$, this is equivalent to showing that the only solution to $A\boldsymbol{x} = \boldsymbol{0}$ is $\boldsymbol{x} = \boldsymbol{0}$.

□

Corollary. *S is linearly independent if and only if* $\operatorname{rref}(A)$ *has a pivot in every column. Furthermore, if $n > m$ (i.e., the number of vectors n in S is strictly greater than the dimension m of $\mathbb{R}^m$) then S cannot be linearly independent.*

Proof. $A\boldsymbol{x} = \boldsymbol{0}$ has a unique solution if and only if $\operatorname{rref}(A)$ does not yield free variables—that is, if there is a pivot in every column of $\operatorname{rref}(A)$. In the case when $n > m$, we must have some non-pivot columns in $\operatorname{rref}(A)$ since A has more columns than rows. Using the previous theorem completes the proof.

□

1.2.2 Maximal linearly independent subset

Now suppose we have a set of vectors which are not linearly independent. Can we discard some of these vectors in order to end up with a linearly independent set? This leads to the notion of a maximal linearly independent subset.

Definition. *Let $S \subseteq \mathcal{V}$. We say that B is a* ***maximal linearly independent subset*** *of S if these two properties hold:*

i. B is a linearly independent subset of S

ii. If $B \subset C$ with $C \neq B$, then C is linearly dependent.

Given $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subseteq \mathbb{R}^m$, how we compute a maximal linearly independent subset B? We use a system of equations to find an answer. Suppose $A\boldsymbol{x} = \boldsymbol{0}$, where A is the matrix with elements of S as its columns as before. We know that this equation has nontrivial solutions $\boldsymbol{x}$ if $\operatorname{rref}(A)$ has columns with no pivot (since this yields free variables). Since we want the zero vector to be the only solution, we must exclude all the non-pivot columns of A in order for this to hold true. This is equivalent to writing $\tilde{A}\tilde{\boldsymbol{x}} = \boldsymbol{0}$, where $\tilde{A}$ is the smaller matrix consisting of only the pivot columns of A. To summarize, we have proved:

Theorem 6. *Let A be the matrix which has $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\} \subseteq \mathbb{R}^m$ as its columns. Then a maximal linearly independent subset of S is given by the pivot columns of A.*

Corollary. *B is a maximal linearly independent subset of S if and only if it is a minimal spanning subset of S.*

Example. Suppose $S = \left\{[1,\ 1]^T,\ [-2,\ -1]^T,\ [1,\ 0]^T,\ [0,\ 1]^T\right\}$ as before. The pivot columns $B = \left\{[1, 1]^T, [-2, -1]^T\right\}$ of A comprise a maximal linearly independent subset of S. Again, maximal linearly independent subsets are not unique—in this example, the set $\tilde{B} = \left\{[1, 0]^T, [0, 1]^T\right\}$ is also such a set.

1.3 Basis and dimension

In this section, we study subsets $B \subseteq \mathcal{V}$ which span all of $\mathcal{V}$ and are also linearly independent. By the previous results this implies that *every* vector in $\mathcal{V}$ has a *unique* representation as a linear combination of B. This is important since having such a B allows us to specify any element of the vector space simply in terms of a set of scalars.

1.3.1 Definitions and basic results

Definition. *$B \subseteq \mathcal{V}$ is a **basis** for $\mathcal{V}$ if and only if the following are true:*

i. B spans $\mathcal{V}$

ii. B is linearly independent.

The prototypical example of a basis is the set of standard basis vectors $B = \{\boldsymbol{e}_1, ..., \boldsymbol{e}_n\}$ in $\mathbb{R}^n$, where $\boldsymbol{e}_k$ is the vector which has k^{th} component equal to 1 and all other components equal to 0. It is easy to check that this set of vectors is linearly independent and that any vector in $\mathbb{R}^n$ can be written as a linear combination of them.

We begin with a technical lemma that is proven in the book (see p. 257-259 of Andrilli and Hecker):

Lemma 7. *Suppose S is a finite set that spans $\mathcal{V}$. If $T \subseteq \mathcal{V}$ is linearly independent, then T must be finite and $|T| \leq |S|$.*

As we now show, this implies that if a vector space has a finite basis, then all other bases of this space are also finite and have the same number of elements.

Theorem 8. *Let B_1 and B_2 be bases of $\mathcal{V}$ such that B_1 is finite. Then B_2 is finite and $|B_1| = |B_2|$.*

Proof. Since B_1 is a basis, it spans $\mathcal{V}$. In addition, B_2 is linearly independent since it is a basis. By the previous lemma, finiteness of B_1 implies finiteness of B_2 and that $|B_2| \leq |B_1|$. Since we have shown B_2 is finite, we can reverse the roles of B_1 and B_2 and apply the lemma again to find $|B_1| \leq |B_2|$. Therefore, $|B_1| = |B_2|$ and the proof is complete.

□

This theorem gives us an unambiguous notion of dimension of a vector space. Roughly, the dimension of a vector space corresponds to the number of "independent directions" in which we can move. A precise characterization of this notion as follows:

Definition. *If $\mathcal{V}$ has a finite basis B we say that $\mathcal{V}$ is **finite-dimensional** and define the **dimension** of $\mathcal{V}$ to be* $\dim(\mathcal{V}) = |B|$. *If $\mathcal{V}$ has no finite basis that we say that $\mathcal{V}$ is **infinite-dimensional**.*

The next result shows that for a finite-dimensional space $\mathcal{V}$ with $\dim(\mathcal{V}) = n$, any set that spans $\mathcal{V}$ must have *at least* n elements, and any set that is linearly independent must have *at most* n elements. The boundary case for each exactly corresponds to the set being a basis, which must both span $\mathcal{V}$ and be linearly independent.

Theorem 9. *Suppose $\mathcal{V}$ is finite-dimensional.*

1. *If S is a finite subset that spans $\mathcal{V}$ then $|S| \geq \dim(\mathcal{V})$. Moreover, $|S| = \dim(\mathcal{V})$ if and only if S is a basis for $\mathcal{V}$.*
2. *If T is a linearly independent subset of $\mathcal{V}$ then $|T| \leq \dim(\mathcal{V})$. Moreover, $|T| = \dim(\mathcal{V})$ if and only if T is a basis for $\mathcal{V}$.*

Proof. Suppose B is a basis for $\mathcal{V}$ with $|B| = \dim(\mathcal{V}) = n$. Applying the technical lemma with $T = B$ for statement (1) and $S = B$ for statement (2) proves the the inequalities given above. Now we must confirm that equality holds in each statement if and only if the set in question is actually a basis.

For (1), assume $|S| = \dim(\mathcal{V}) = n$. We prove that S is a basis by contradiction. Since S spans $\mathcal{V}$, if it is not a basis then it must be linearly independent. Therefore, there is some $\boldsymbol{v} \in S$ such that $\boldsymbol{v} \in \operatorname{span}(S - \{\boldsymbol{v}\})$. So, $\boldsymbol{v}$ is redundant and we have that $S - \{\boldsymbol{v}\}$ spans $\mathcal{V}$. But this is a contradiction since $|S - \{\boldsymbol{v}\}| = n - 1 < \dim(\mathcal{V})$, which we just proved cannot be true since every spanning set must have size greater than or equal to $\dim(\mathcal{V})$! For the converse, if S is a basis for $\mathcal{V}$ then we know that S is finite and $|S| = \dim(\mathcal{V})$ by the previous theorem.

For (2), assume $|T| = \dim(\mathcal{V}) = n$. We prove that T is a basis by contradiction. Since T is linearly independent, if it is not a basis then it must not span $\mathcal{V}$. Therefore, there is some vector $\boldsymbol{v} \in \mathcal{V}$ such that $T \cup \{\boldsymbol{v}\}$ is linearly independent. But this is a contradiction since $|T \cup \{\boldsymbol{v}\}| = n + 1 > \dim(\mathcal{V})$, which cannot be true as we have proven that every linearly independent set has size smaller than or equal to $\dim(\mathcal{V})$. Finally, for the converse we know that if T is a basis for $\mathcal{V}$ then $|T| = \dim(\mathcal{V})$ by the previous theorem.

□

1.3.2 Constructing bases from spanning sets and linearly independent sets

The next result is essentially a restatement of the previous theorem, except that it does not explicitly use the dimension of $\mathcal{V}$ (and it also holds for infinite-dimensional spaces). Since it is similar in spirit to the previous theorem we do not provide a proof.

Theorem 10. *Suppose S spans $\mathcal{V}$. Then a minimal spanning set (equivalently, a maximal linearly independent set) $B \subseteq S$ is a basis for $\mathcal{V}$.*

This implies that if we have a set that spans $\mathcal{V}$ then we can always find a basis for $\mathcal{V}$ contained within it by throwing away redundant vectors. In addition, we can show that given a linearly independent set, we can always find a basis for $\mathcal{V}$ that contains it by adding additional vectors. To summarize:

Theorem 11. *Suppose $\mathcal{V}$ is finite-dimensional.*

1. *If S spans $\mathcal{V}$, there is some basis B for $\mathcal{V}$ such that $B \subseteq S$.*
2. *If T is linearly independent, there is some basis B for $\mathcal{V}$ such that $T \subseteq B$.*

Proof. For statement (1), simply take B to be a minimal spanning set of S. Then $B \subseteq S$ trivially and B is a basis. For statement (2), let C be any set that spans $\mathcal{V}$ and consider the finite set $T \cup C$. Order the elements of $T \cup C$ so that the elements of T appear first. Then, we can iteratively check for linear independence of larger and larger subsets to find a maximal linearly independent subset B that contains T (since T is itself linearly independent). Then B is a basis and we are done.

□

The theorem above gives a computational method for find a basis in $\mathbb{R}^m$ that contains a given spanning set, or is contained in a given linearly independent set:

- *Case 1 (shrinkage)*: Given S that spans $\mathbb{R}^m$ but is not necessarily linearly independent, find a minimal spanning subset B. Then B is a basis that is contained in S.
- *Case 2 (enlargement)*: Given a linearly independent set T that does not necessarily span $\mathbb{R}^m$, add on a spanning set C for $\mathcal{V}$ and then find a maximal linearly independent subset B that contains T (to do this, first put the vectors in T as the columns of a matrix A before adding vectors in C to the right of it). Then B is a basis.

Example. Suppose $\mathcal{V} = \mathbb{R}^2$ and $S = \{[1,1]^T, [-2,-1]^T, [1,0]^T, [0,1]^T\}$. Then S spans $\mathbb{R}^2$ and by the computation done earlier, the minimal spanning subset $B = \{[1,1]^T, [-2,-1]^T\}$ is a basis for $\mathbb{R}^2$ which is contained in S.

Example. Suppose $T = \{[4,4]^T\}$, which is obviously linearly independent. To find a basis that contains T, we add to it any spanning set C of $\mathbb{R}^2$—for example, $C = \{[1,1]^T, [-2,-1]^T, [1,0]^T, [0,1]^T\}$. Putting the vectors in $T \cup C$ as the columns of a matrix $A = \begin{bmatrix} 4 & 1 & -2 & 1 & 0 \\ 4 & 1 & -1 & 0 & 1 \end{bmatrix}$ (where the vectors in T have been added first), we have $\operatorname{rref}(A) = \begin{bmatrix} 1 & \frac{1}{4} & 0 & -\frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 1 & -1 & 1 \end{bmatrix}$. Since the first and third columns of A are the pivot columns, we find that a maximal linearly independent subset of $T \cup C$ that contains T is $B = \{[4,4]^T, [-2,-1]^T\}$.

1.3.3 Dimension of a subspace

We conclude our discussion with a theorem concerning subspaces:

Theorem 12. *Suppose $\mathcal{V}$ is finite-dimensional and $\mathcal{W}$ is a subspace of $\mathcal{V}$. Then $\mathcal{W}$ is finite-dimensional with* $\dim(\mathcal{W}) \leq \dim(\mathcal{V})$*. Moreover,* $\dim(\mathcal{W}) = \dim(\mathcal{V})$ *if and only if $\mathcal{W} = \mathcal{V}$.*

Intuitively, this should make sense. Since a subspace is no larger than the vector space in which it resides, we should expect that its dimension cannot be larger either. The only subtlety in the proof of this theorem is to demonstrate that $\mathcal{W}$ actually has a basis, which we will not show here.

2 Coordinatization

We can study general finite-dimensional vector spaces $\mathcal{V}$ using the same tools we used for $\mathbb{R}^n$, so long as we can find a way to relate $\mathcal{V}$ to $\mathbb{R}^n$. This process is known as coordinatization.

2.1 Representation of vectors in a basis

Suppose $B = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ is a basis of an n-dimensional vector space $\mathcal{V}$. We claim that for any $\boldsymbol{v} \in \mathcal{V}$, there is a *unique* way of writing $\boldsymbol{v}$ as a linear combination of B. To justify this, if $\boldsymbol{v} = \sum_{i=1}^n a_i \boldsymbol{b}_i$ for some set of scalars $a_1, \ldots, a_n$ and we can also write $\boldsymbol{v} = \sum_{i=1}^n c_i \boldsymbol{b}_i$ for some set of scalars $c_1, \ldots, c_n$, then $\sum_{i=1}^n (a_i - c_i)\boldsymbol{b}_i = \mathbf{0}$ and linear independence of B implies $a_i = c_i$ for every i. Knowing this, we have the following:

Definition. *Given a basis $B = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$, suppose $\boldsymbol{v} = \sum_{i=1}^n a_i \boldsymbol{b}_i$. Then the* ***coordinates*** *$[\boldsymbol{v}]_B \in \mathbb{R}^n$ of $\boldsymbol{v}$ in B is the real-valued vector*

$$[\boldsymbol{v}]_B = [a_1, \ldots, a_n]^T.$$

Note that the coordinatization of the i^{th} basis vector $\boldsymbol{b}_i$ is $[\boldsymbol{b}_i]_B = \boldsymbol{e}_i = [0, \ldots, 0, 1, 0, \ldots, 0]^T$, the i^{th} standard basis vector in $\mathbb{R}^n$. Furthermore, coordinatization preserves the operations of vector addition and scalar multiplication, and is therefore an *isomorphism* from $\mathcal{V}$ to $\mathbb{R}^n$:

Theorem 13.

i. *Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$ and $a \in \mathbb{R}$. Then $[\boldsymbol{x} + \boldsymbol{y}]_B = [\boldsymbol{x}]_B + [\boldsymbol{y}]_B$ and $[a\boldsymbol{x}]_B = a[\boldsymbol{x}]_B$.*

ii. *If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in \mathcal{V}$ and $\boldsymbol{w} = \sum_{i=1}^m a_i \boldsymbol{v}_i$, then $[\boldsymbol{w}]_B = \sum_{i=1}^m a_i [\boldsymbol{v}_i]_B$. Therefore, $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ spans $\mathcal{V}$ if and only if $\{[\boldsymbol{v}_1]_B, \ldots, [\boldsymbol{v}_m]_B\}$ spans $\mathbb{R}^n$.*

iii. *$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ are linearly independent if and only if $\{[\boldsymbol{v}_1]_B, \ldots, [\boldsymbol{v}_m]_B\}$ are linearly independent.*

Since the proof of these statements is straightforward, we do not provide them here. The main utility of this result is that it allows us to compute the span and check for linear independence of vectors in an abstract space $\mathcal{V}$ using the methods we discussed earlier for $\mathbb{R}^n$.

2.2 Change of basis

Suppose $B = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ and $C = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\}$ are two bases of $\mathcal{V}$. If we know the coordinates $[\boldsymbol{v}]_B$ of a vector $\boldsymbol{v} \in \mathcal{V}$ in basis B, can we use this to find the coordinates $[\boldsymbol{v}]_C$ in basis C?

Theorem 14. *There is a unique $n \times n$ matrix P_{CB} such that $[\boldsymbol{v}]_C = P_{CB}[\boldsymbol{v}]_B$ for all $\boldsymbol{v} \in \mathcal{V}$. Furthermore, the i^{th} column of P_{CB} equals $[\boldsymbol{b}_i]_C$, so that*

$$P_{CB} = \left[\begin{array}{ccc} [\boldsymbol{b}_1]_C & \cdots & [\boldsymbol{b}_n]_C \end{array}\right].$$

Proof. Suppose $\boldsymbol{v} = \sum_{i=1}^{n} a_i \boldsymbol{b}_i$, so that $[\boldsymbol{v}]_B = [a_1, \ldots, a_n]^T$. Then by direct calculation,

$$[\boldsymbol{v}]_C = \left[\sum_{i=1}^{n} a_i \boldsymbol{b}_i\right]_C = \sum_{i=1}^{n} a_i [\boldsymbol{b}_i]_C = \left[\begin{array}{ccc} [\boldsymbol{b}_1]_C & \cdots & [\boldsymbol{b}_n]_C \end{array}\right] \left[\begin{array}{c} a_1 \\ \vdots \\ a_n \end{array}\right] = P_{CB} [\boldsymbol{v}]_B.$$

To check that P_{CB} is the unique matrix which satisfies this relation, suppose that there is an $n \times n$ matrix A such that $[\boldsymbol{v}]_C = A[\boldsymbol{v}]_B$ for all $\boldsymbol{v} \in \mathcal{V}$. Then $[\boldsymbol{b}_i]_C = A[\boldsymbol{b}_i]_B = A\boldsymbol{e}_i = i^{\text{th}}$ column of A, so we must have $A = P_{CB}$.

□

We call P_{CB} the **change of basis matrix** (or **transition matrix**) from B to C. Note that to keep track of which basis we are starting from and moving to, the subscripts should be read from right to left.

Corollary. *If B and C are bases of $\mathcal{V}$, then P_{CB} is invertible and the change of basis matrix P_{BC} from C to B satisfies $P_{BC} = P_{CB}^{-1}$.*

Proof. We only need to show that $P_{BC} P_{CB} = I_n$. Let $\boldsymbol{y} \in \mathbb{R}^n$. Since $[\cdot]_B$ is an isomorphism, there is some $\boldsymbol{x} \in \mathcal{V}$ such that $[\boldsymbol{x}]_B = \boldsymbol{y}$. Therefore,

$$P_{BC} P_{CB} \boldsymbol{y} = P_{BC} P_{CB} [\boldsymbol{x}]_B = P_{BC} [\boldsymbol{x}]_C = [\boldsymbol{x}]_B = \boldsymbol{y}$$

for every $\boldsymbol{y} \in \mathbb{R}^n$ and $P_{BC} P_{CB}$ is the identity matrix.

□

In practice, given a standard basis B and another basis C it is often difficult to compute P_{CB} but trivial to find P_{BC}. Using the corollary above gives us a straightforward way of generating P_{CB} by finding the inverse of P_{BC}. To summarize, the procedure to compute $[\boldsymbol{v}]_C$ is:

1. Find $[\boldsymbol{v}]_B$ (this step is usually trivial).
2. Find P_{BC} by computing $[\boldsymbol{c}_i]_B$ for each i.
3. Invert P_{BC} to generate $P_{CB} = P_{BC}^{-1}$.
4. Evaluate $[\boldsymbol{v}]_C = P_{CB} [\boldsymbol{v}]_B$.

Remark. This procedure is essentially what is described in Andrilli & Hecker, p. 284-288, except that we are assuming from the outset that $\boldsymbol{v} \in \mathcal{V} = \operatorname{span}(B) = \operatorname{span}(C)$. In the book, the procedure is described using augmented matrices, which in our discussion is used to find P_{CB} by inverting P_{BC}. Finally, note that the textbook uses P to denote the change of basis matrix P_{CB}, and P^{-1} to denote P_{BC}.

Example. Let $B = \left\{ \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array}\right], \left[\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array}\right], \left[\begin{array}{cc} 0 & 0 \\ 1 & 0 \end{array}\right], \left[\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array}\right] \right\}$ be the standard basis in $\mathcal{M}_{22}$. It can be checked that $C = \left\{ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right], \left[\begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array}\right], \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right], \left[\begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array}\right] \right\}$ is also a basis for $\mathcal{M}_{22}$ (how?). Defining $\boldsymbol{v} = \left[\begin{array}{cc} 1 & 2 \\ 4 & 3 \end{array}\right] \in \mathcal{M}_{22}$, we easily find that $[\boldsymbol{v}]_B = [1, 2, 4, 3]^T$. What is $[\boldsymbol{v}]_C$?

Unlike when we were working with basis B, finding the coordinates in C is nontrivial since it is not clear how to write $\boldsymbol{v}$ as a linear combination of vectors in C. To do this we will use the change of basis matrix $P_{CB} = P_{BC}^{-1}$. Since $[\boldsymbol{c}_1]_B = [1, 0, 0, 1]^T$, $[\boldsymbol{c}_2]_B = [1, 0, 0, 1]^T$, $[\boldsymbol{c}_3]_B = [0, 1, 1, 0]^T$, and $[\boldsymbol{c}_4]_B = [0, 1, -1, 0]^T$, we find

$$P_{BC} = \left[\begin{array}{ccc} [\boldsymbol{c}_1]_B & \cdots & [\boldsymbol{c}_4]_B \end{array}\right] = \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \\ 1 & -1 & 0 & 0 \end{array}\right].$$

Inverting this matrix implies that

$$P_{CB} = P_{BC}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

so

$$[\boldsymbol{v}]_C = P_{CB}[\boldsymbol{v}]_B = [2,-1,3,-1]^T.$$

That is, $\boldsymbol{v} = 2\boldsymbol{c}_1 - \boldsymbol{c}_2 + 3\boldsymbol{c}_3 - \boldsymbol{c}_4$, which can be verified directly.

We conclude this section with a general result that encompasses the ones discussed earlier. It provides us a criterion for checking, given a basis B of a subspace $\mathcal{W} = \operatorname{span}(B)$, whether a set C of the same size is also a basis of $\mathcal{W}$ (note that if $|C| \neq |B|$ then C cannot possibly be a basis):

Theorem 15. *Suppose E is a basis for $\mathcal{V}$, and let B be a basis for a nontrivial k-dimensional subspace $\mathcal{W}$ of $\mathcal{V}$. Let C be a set such that $|C| = |B|$, and define the $n \times k$ matrices*

$$P_{EB} = \begin{bmatrix} [\boldsymbol{b}_1]_E & \cdots & [\boldsymbol{b}_k]_E \end{bmatrix}, \qquad P_{EC} = \begin{bmatrix} [\boldsymbol{c}_1]_E & \cdots & [\boldsymbol{c}_k]_E \end{bmatrix}.$$

Then C is a basis for $\mathcal{W}$ if and only if

$$\operatorname{rref}([P_{EC}|P_{EB}]) = \left[\begin{array}{c|c} I_k & R \\ \hline O_{n-k,k} & O_{n-k,k} \end{array}\right]$$

where I_k is the $k \times k$ identity matrix, $O_{n-k,k}$ is the $(n-k) \times k$ zero matrix, and R is some $k \times k$ matrix. Furthermore, if C is indeed a basis then R is the change of basis matrix from B to C—i.e., $P_{CB} = R$.

The usefulness of the theorem is that it allows us to find the change of basis matrix P_{CB} when neither B nor C are simple or standard, and even in the case when we do not explicitly know the span of B or C. No proof is given since this result is similar in spirit to those discussed earlier, but more general and stated in a slightly different manner. To see this, consider the case when $\mathcal{W} = \mathcal{V}$, $E = B$, and if C is assumed to be a basis of $\mathcal{W}$. Then, $P_{EB} = P_{BB} = I_n$ and $P_{EC} = P_{BC}$, so the conclusion of the theorem is that

$$\operatorname{rref}([P_{BC}|I_n]) = [I_n|P_{CB}],$$

i.e., that $P_{CB} = P_{BC}^{-1}$.

Example. Let $\mathcal{V} = \mathcal{P}_4$ and consider the basis $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$ of $\mathcal{W} = \operatorname{span}(B)$, where

$$\begin{aligned} \boldsymbol{b}_1 &= 6x^4 + 20x^3 + 7x^2 + 19x - 4 \\ \boldsymbol{b}_2 &= x^4 + 5x^3 + 7x^2 - x + 6 \\ \boldsymbol{b}_3 &= 5x^3 + 17x^2 - 10x + 19. \end{aligned}$$

Define $C = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3\}$, where

$$\begin{aligned} \boldsymbol{c}_1 &= x^4 + 3x^3 + 4x - 2 \\ \boldsymbol{c}_2 &= 2x^4 + 7x^3 + 4x^2 + 3x + 1 \\ \boldsymbol{c}_3 &= 2x^4 + 5x^3 - 3x^2 + 8x - 7. \end{aligned}$$

Is C also a basis of $\mathcal{W}$, and if so, what is the change of basis matrix from B to C?

To begin, let us start by writing B and C in coordinates of the standard basis $E = \{1, x, x^2, x^3, x^4\}$ of $\mathcal{P}_4$. That is,

$$P_{EB} = \begin{bmatrix} [\boldsymbol{b}_1]_E & \cdots & [\boldsymbol{b}_3]_E \end{bmatrix} = \begin{bmatrix} -4 & 6 & 19 \\ 19 & -1 & -10 \\ 7 & 7 & 17 \\ 20 & 5 & 5 \\ 6 & 1 & 0 \end{bmatrix}$$

and

$$P_{EC}=\left[\begin{array}{ccc}[\boldsymbol{c}_1]_E & \cdots & [\boldsymbol{c}_3]_E\end{array}\right]=\left[\begin{array}{ccc}-2&1&-7\\4&3&8\\0&4&-3\\3&7&5\\1&2&2\end{array}\right].$$

We find that

$$\operatorname{rref}([P_{EC}|P_{EB}])=\left[\begin{array}{ccc|ccc}1&0&0&6&1&2\\0&1&0&1&1&2\\0&0&1&-1&-1&-3\\\hline 0&0&0&0&0&0\\0&0&0&0&0&0\end{array}\right]=\left[\begin{array}{c|c}I_3&P_{CB}\\\hline O_{2,3}&O_{2,3}\end{array}\right]$$

so C is a basis for $\mathcal{W}$ and $P_{CB}=\left[\begin{array}{ccc}6&1&2\\1&1&2\\-1&-1&-3\end{array}\right]$.

Example. Suppose $\mathcal{V}=\mathcal{S}_{2,2}$, the vector space of 2×2 *symmetric* matrices. Let $B=\left\{\left[\begin{array}{cc}2&-3\\-3&1\end{array}\right],\left[\begin{array}{cc}1&5\\5&0\end{array}\right]\right\}$. It is easy to check that B is linearly independent, so it is a basis for $\mathcal{W}=\operatorname{span}(B)$. Now consider the set $C=\left\{\left[\begin{array}{cc}0&2\\2&-4\end{array}\right],\left[\begin{array}{cc}-1&6\\6&3\end{array}\right]\right\}$. Is C a basis of $\mathcal{W}$?

To begin, it can be checked directly that $E=\left\{\left[\begin{array}{cc}1&0\\0&0\end{array}\right],\left[\begin{array}{cc}0&1\\1&0\end{array}\right],\left[\begin{array}{cc}0&0\\0&1\end{array}\right]\right\}$ is a basis of $\mathcal{S}_{22}$ (how?). Then

$$P_{EB}=\left[\begin{array}{cc}[\boldsymbol{b}_1]_E & [\boldsymbol{b}_2]_E\end{array}\right]=\left[\begin{array}{cc}2&1\\-3&5\\1&0\end{array}\right],\qquad P_{EC}=\left[\begin{array}{cc}[\boldsymbol{c}_1]_E & [\boldsymbol{c}_2]_E\end{array}\right]=\left[\begin{array}{cc}0&-1\\2&6\\-4&3\end{array}\right].$$

Since

$$\operatorname{rref}([P_{EC}|P_{EB}])=\left[\begin{array}{cc|cc}1&0&0&1\\0&1&0&1\\\hline 0&0&1&1\end{array}\right]\neq\left[\begin{array}{c|c}I_2&R\\\hline O_{1,2}&O_{1,2}\end{array}\right],$$

C is not a basis of $\mathcal{W}$.