An alternative approach to (the teaching of) rank and dimension

Carl de Boor

Preliminaries

While the material in this note is meant to be taught in a first course in Linear Algebra, it is written for those teaching that course (rather than those taking it). Since maps play an essential role in this material, I would assume such a course to begin with a detailed introduction to maps, to cover such items as: the notation $f : X \to Y : x \mapsto f(x)$ and the corresponding abbreviations $\operatorname{dom} f := X, \operatorname{tar} f := Y, \operatorname{ran} f := f(X)$ for the domain, the target, and the range of the map f, the identity map id_X from X to X, the composition of maps and the fact that map composition is associative, and the basic properties: 1-1, onto, invertible. Undergraduates often have trouble with these basic concepts, and Linear Algebra offers an ideal opportunity to provide enough work with maps to overcome this trouble.

The discussion below further assumes as known the definition: (i) of a vector space over some field $\mathbb{F}$ (even if it is only the real field); (ii) of the coordinate spaces $\mathbb{F}^n := \{(f_1, f_2, \ldots, f_n) : f_j \in \mathbb{F}\}$, $n = 0, 1, 2, \ldots$ and of the space $\mathbb{F}^{m\times n}$ of $m \times n$-matrices; (iii) of a linear map; and (iv) of the space $L(X, Y)$ of all linear maps with domain X and target Y. It also assumes as known that the composition of linear maps is linear, and that a linear map is 1-1 if and only if its nullspace is trivial.

Linear maps from coordinate space

Among the linear maps, those with domain one of the coordinate spaces $\mathbb{F}^n$ play a special role. Any such map $V \in L(\mathbb{F}^n, Y)$ is characterized by its **columns** $v^j := Ve^j$, $j = 1, \ldots, n$, in the sense that

$$Va = v^1a_1 + v^2a_2 + \cdots + v^na_n,$$

with e^j the jth unit vector. (Throughout, a superscript denotes a particular term in a sequence of vectors, while a subscript denotes a particular entry in a sequence of scalars.) It is at times convenient to display the characterizing sequence explicitly by writing $[v^1, v^2, \ldots, v^n]$ for the linear map given by the rule

$$[v^1, v^2, \ldots, v^n] : \mathbb{F}^n \to Y : a \mapsto v^1a_1 + v^2a_2 + \cdots + v^na_n.$$

I call the number of columns of such V its **order** and denote it by $\#V$. I also call any such map a **column map** (for want of a better term).

When $Y = \mathbb{F}^m$, $V = [v^1, v^2, \ldots, v^n]$ becomes (or is represented by) the $m \times n$-matrix with columns $v^1, v^2, \ldots, v^n$. Thus, $V \in \mathbb{F}^{m\times n}$, and the action of V on some $c \in \mathbb{F}^n$ can be described by

$$(Vc)_i = \sum_k v^k_i c_k = \sum_k V(i, k)c_k, \qquad \text{all } i.$$

The easily verifiable fact that $A[v^1, v^2, \ldots, v^n] = [Av^1, Av^2, \ldots, Av^n]$ for any $A \in L(X, Y)$ is the basis for the peculiar way in which we define the product of two matrices. It is at this point that it becomes appropriate to do some matrix algebra. Of particular interest for the material below is the fact, easily verified, that for any two compatible matrices A and B, $(AB)' = B'A'$, with A' the transpose or the conjugate transpose of A. The only other matrix result needed is the central result from elimination that a homogeneous linear system with more unknowns than equations has nontrivial solutions.

(1)Theorem. *If* $A \in \mathbb{F}^{m\times n}$ *with* $m < n$, *then* $\operatorname{null} A \neq \{0\}$.

Rank

A factorization $A = V\Lambda$ of $A \in L(X,Y)$ into $\Lambda \in L(X,\mathbb{F}^n)$ and $V \in L(\mathbb{F}^n, Y)$ is essential for any kind of computation with A, as we can only compute in coordinate space. We call the integer $n = \#V$ the **order** of the factorization $A = V\Lambda$. The smaller the order n, the simpler the calculations. Hence the smallest possible order is of particular interest. We call this smallest possible order the **rank** of A. In symbols:

$$\text{rank}\, A := \min\{\#V : A = V\Lambda\}. \tag{2}$$

We say that $A \in L(X,Y)$ is **of finite rank** if it has such a factorization, since then the collection of all such factorizations is not empty, and all have nonnegative order, therefore there is a **minimal factorization**, i.e., a factorization of minimal order.

Any linear map from or to a coordinate space is of finite rank: If $A \in L(\mathbb{F}^n, Y)$, then $A = A\text{id}_n$ is a (trivial) factorization, hence $\text{rank}\, A \le n$. Similarly, any $A \in L(X, \mathbb{F}^m)$ is of finite rank, since such a map has the (trivial) factorization $A = \text{id}_m A$. In particular, $\text{rank}\, A \le \min\{m,n\}$ for any $A \in \mathbb{F}^{m\times n}$. Also, with A' the transpose or the conjugate transpose of $A \in \mathbb{F}^{m\times n}$, $\text{rank}\, A' = \text{rank}\, A$ for any $A \in \mathbb{F}^{m\times n}$, since $A = V\Lambda$ if and only if $A' = \Lambda' V'$.

(3)Theorem. *The factorization $A = V\Lambda$ for $A \in L(X,Y)$ is minimal if and only if V is 1-1 and* $\text{ran}\, V = \text{ran}\, A$.

The proof of this Theorem is given in a sequence of Lemmas of independent interest.

Proofs

(4)Lemma. *If $V \in L(\mathbb{F}^n, Y)$ is 1-1 and $W \in L(\mathbb{F}^m, Y)$ is onto, then $n \le m$.*

Proof: Since W is onto, we can find, for each column v^j of V, some m-vector c^j so that $v^j = Wc^j$. This shows that $V = WC$, with $C := [c^1, c^2, \ldots, c^n] \in \mathbb{F}^{m\times n}$. If now $m < n$, then C would not be 1-1 (by (1)Theorem), hence V would not be 1-1, contrary to our assumption. □

(5)Corollary. *If $A = V\Lambda$ with the column map V 1-1 and* $\text{ran}\, V = \text{ran}\, A$, *then* $\text{rank}\, A = \#V$.

Proof: Since $A = V\Lambda$, we have $\text{rank}\, A \le \#V$. On the other hand, whenever $A = WM$ for some column map W, then necessarily $\text{ran}\, V = \text{ran}\, A \subset \text{ran}\, W$, hence $\text{dom}\, V \to \text{ran}\, W : a \mapsto Va$ is a 1-1 column map, therefore $\#V \le \#W$ by (4)Lemma. Consequently, $\#V \le \min\{\#W : A = WM\} = \text{rank}\, A$. □

(6)Lemma. *If $A = V\Lambda$ is minimal, then V is 1-1.*

Proof: If $V \in L(\mathbb{F}^n, Y)$ is not 1-1, then its nullspace is not trivial, i.e., there is some $a \in \mathbb{F}^n \backslash 0$ for which $Va = 0$. Since $a \ne 0$, there must be some j so that $a_j \ne 0$. In order to avoid some funny indexing or reordering, assume that, in fact, $a_n \ne 0$. Since $0 = Va = v^1 a_1 + v^2 a_2 + \cdots + v^n a_n$, we conclude that

$$v^n = v^1 b_1 + v^2 b_2 + \cdots + v^{n-1} b_{n-1},$$

with b the $(n-1)$-vector with entries $b_i := -a_i/a_n$, all i. This implies that $V = V_1 M$, with

$$V_1 := [v^1, v^2, \ldots, v^{n-1}]$$

and $M := [e^1, e^2, \ldots, e^{n-1}, b] \in \mathbb{F}^{(n-1)\times n}$, hence $A = V_1(M\Lambda)$ provides a factorization for A of order less than n, showing that $A = V\Lambda$ is not minimal. □

The proof establishes the following slightly stronger result of use in the construction of minimal factorizations.

(7)Corollary. *If the column map V is not 1-1, then $V = V_1 M$, with V_1 made up of all but one of the columns of V.*

Repeated application of this Corollary provides, for any $V \in L(\mathbb{F}^n, Y)$, a factorization $V = V_s M_s$, with V_s 1-1 and made up entirely of columns of V. The last fact implies that $\operatorname{ran} V_s \subset \operatorname{ran} V$, hence $\operatorname{ran} V_s = \operatorname{ran} V$, therefore $V = V_s M_s$ is a minimal factorization for V, by (5)Corollary. I record this result, for later reference.

(8)Corollary. *Every column map V has a minimal factorization $V = V_s M_s$, with every column of V_s also a column of V.*

At this point, all the assertions of (3)Theorem have been proved, except for the implication that, for a minimal factorization $A = V\Lambda$, necessarily $\operatorname{ran} V = \operatorname{ran} A$. This implication is trivial in case A is onto, hence we are entitled to use the Theorem in the sequel under this additional assumption. To complete the proof, we need the concept of **dimension**.

Basis and dimension

We call any invertible linear map $V \in L(\mathbb{F}^n, Y)$ a **basis** for Y, and call the n-vector $V^{-1}y$ the **coordinates** of $y \in Y$ with respect to the basis V.

It is customary to reserve the word 'basis' for the sequence $v^1, v^2, \ldots, v^n$ of columns of V and not even to mention the map $[v^1, v^2, \ldots, v^n]$. Further, it is customary to say that $v^1, v^2, \ldots, v^n$ is *linearly independent* (*spanning* for Y) when V is 1-1 (onto), and to call any element of $\operatorname{ran} V$ a *linear combination* of the terms of the sequence $v^1, v^2, \ldots, v^n$. The reason for this particular usage is not clear, given that the simple and basic terms '1-1', 'onto', '$\operatorname{ran} V$' are available.

We conclude from (8)Corollary that any onto map $W \in L(\mathbb{F}^m, Y)$ has a minimal factorization $W = V\Lambda$, with every column of V a column of W. The factorization being minimal, V is necessarily a basis for Y, since V is necessarily onto, while V must be 1-1 by (6)Lemma. This proves

(9)Proposition. *Every column map can be 'thinned' to a basis for its range.*

If we follow custom and call Y **finitely generated** in case it is the range of a column map, then we have

(10)Theorem. *A finitely generated vector space has a basis.*

The identity map id_Y on the vector space Y is trivially onto. Hence (3)Theorem as proved so far provides the conclusion that $\mathrm{id}_Y = V\Lambda$ is minimal if and only if V is invertible, in which case necessarily $V^{-1} = \Lambda$. This proves

(11)Proposition. *The factorization* $\mathrm{id}_Y = V\Lambda$ *for the identity map on Y is minimal if and only if* $V^{-1} = \Lambda$.

It follows that the number $\#V$ of columns in any basis V for Y equals the rank of id_Y. This number is called the **dimension** of Y. In formulae:

$$\dim Y := \operatorname{rank} \mathrm{id}_Y. \tag{12}$$

If, in particular, $\mathrm{id}_{\operatorname{ran} A} = WM$ is minimal (i.e., if W is a basis for $\operatorname{ran} A$), then $A = W(MA)$ is a minimal factorization for A by (3)Theorem, hence

(13)Corollary. $\operatorname{rank} A = \dim \operatorname{ran} A$.

In particular,

$$\dim \operatorname{ran} V \le \#V \tag{14}$$

for any column map V.

Since $\operatorname{rank} A = \operatorname{rank} A'$ for any matrix A (as observed earlier), we obtain the important

(15)Proposition. *For any* $A \in \mathbb{F}^{m\times n}$, $\dim \operatorname{ran} A = \dim \operatorname{ran} A'$.

Finally, I note that the proof of (3)Theorem can be completed with the aid of the important

(16)Proposition. *If* $X \subseteq Y$ *are two vector spaces, and* $\dim Y < \infty$, *then* $\dim X \le \dim Y$, *with equality if and only if* $X = Y$.

whose proof can be found in the next section. For, we now know (from (13)Corollary) that the minimality of the factorization $A = V\Lambda$ implies that $\#V = \dim \operatorname{ran} A$, while $\dim \operatorname{ran} V \le \#V$ from (14), and $A = V\Lambda$ implies that $\operatorname{ran} A \subset \operatorname{ran} V$; therefore $\operatorname{ran} A = \operatorname{ran} V$ by (16).

Extending to a basis

The proof of (16) and other important results concerning dimension rely on the possibility of extending a 1-1 column map to a basis. The basic fact needed for this is contained in the following

(17)Lemma. *Let* $V \in L(\mathbb{F}^n, Y)$ *and* $y \in Y$. *If* V *is 1-1, then,* $y \notin \operatorname{ran} V$ *if and only if* $[V, y]$ *is 1-1.*

Proof: Take $y \in Y \backslash \operatorname{ran} V$ and consider the equation $Vb + yc =: [V, y](b, c) = 0$. If $c \ne 0$, then we find that $y = V(-b/c) \in \operatorname{ran} V$, contrary to our choice of y. Hence we must have $c = 0$, and therefore already $Vb = 0$ and therefore also $b = 0$ since V is 1-1, hence altogether $(b, c) = 0$, showing that also $[V, y]$ is 1-1.

Conversely, if $y \in \operatorname{ran} V$, then $y = Vb$ for some $b \in \mathbb{F}^n$. But then $[V, y](b, -1) = 0$ and $(b, -1) \ne 0$, hence $[V, y]$ is not 1-1. □

(18)Corollary. *Every 1-1 column map into a finite-dimensional vector space can be extended to a basis (for that space).*

The proof is provided by the following Algorithm.

(19)Algorithm: Given: $A \in L(\mathbb{F}^n, Y)$ 1-1 and $W := [w^1, w^2, \ldots, w^n] \in L(\mathbb{F}^m, Y)$ onto.
Sought: a subsequence $u^1, u^2, \ldots, u^r$ of $w^1, w^2, \ldots, w^n$ for which $V := [A, U]$ is invertible.

Initial Step: Initialize $V := A$.

Loop: For $j = 1, \ldots, n$, if $w^j \notin \operatorname{ran} V$, then set $V := [V, w^j]$.

Output: The column map V which is a basis for Y and contains all the columns of A.

For the **proof** of the claim of the Algorithm, observe that, for the final V, every column of W is in $\operatorname{ran} V$ by construction, hence V is onto, while, by (17)Lemma, V is 1-1.

Note that, by starting with the sole linear map $A \in L(\mathbb{F}^0, Y)$, we recover the earlier result (9) that every onto column map can be thinned to a basis.

The computational task of telling whether or not $w^j \in \operatorname{ran} V$ is best handled by elimination. In fact, it is usually not possible to tell whether or not $w^j \in \operatorname{ran} V$, unless one has in hand some 1-1 linear map $\Lambda : Y \to \mathbb{F}^m$ and can compute with the *matrices* ΛA and ΛW. With these matrices in hand, one would apply elimination to the matrix

$$B := \Lambda[A, W]$$

and thereby obtain a classification of the unknowns into bound and free. The columns of B corresponding to bound unknowns provide a basis for $\operatorname{ran} B$. Since Λ is 1-1, the corresponding columns of $[A, W]$ provide a basis for $\operatorname{ran}[A, W] = Y$. Since elimination picks an unknown as bound if and only if its column is not in the range of the preceding columns, and since A is 1-1, it follows that, in particular, A is part of the resulting basis for Y.

Here, finally, is the **proof** of (16)Proposition. Consider all possible 1-1 column maps V to X. There is at least one, viz. the trivial map $\mathbb{F}^0 \to X$ which maps the sole element of $\mathbb{F}^0$ to 0. Further, $\#V \le \dim Y$ from (4)Lemma, since $X \subseteq Y$. Therefore there is some 1-1 column map V to X with $\#V$ as large as possible. By (17)Lemma, $\operatorname{ran} V = X$ for such maximal V, i.e., such V is a basis for X. In particular, $\dim X = \#V \le \dim Y$.

If now $\dim X = \dim Y$, then $c \mapsto Vc$ is also maximally 1-1 when considered as a map into Y, hence necessarily a basis for Y and so, in particular, $X = Y$. □

As an illustration of the (also notational) ease which the proposed approach provides, here are two more basic results concerning dimensions.

(20)Theorem. *For every linear map A,* $\dim \operatorname{dom} A = \dim \operatorname{ran} A + \dim \operatorname{null} A$.

Proof: Start with a basis U for $\operatorname{null} A$ and use (18)Corollary to extend it to a basis $V := [U, W]$ for $\operatorname{dom} A$. Then $AV = A[U, W] = [AV, AW] = [0, AW]$, hence $\operatorname{ran} A = \operatorname{ran} AW$. Also, AW is 1-1: for, if $AWc = 0$, then $Wc \in \operatorname{null} A$, hence $Wc = Ud$ for some d, therefore $[U, W](-d, c) = 0$, hence $(-d, c) = 0$ since $[U, W]$ is a basis, therefore, finally, $c = 0$. It follows that AW is a basis for $\operatorname{ran} A$, therefore $\dim \operatorname{ran} A = \#AW = \#W = \#V - \#U = \dim \operatorname{dom} A - \dim \operatorname{null} A$. □

(21)Theorem. *If X, Y are vector spaces and X is finite-dimensional, then* $\dim X = \dim Y$ *if and only if there exists an invertible $A \in L(X, Y)$.*

Proof: Let V, W be a basis for X, Y, respectively. If $\dim X = n := \dim Y$, then $\operatorname{dom} V = \operatorname{dom} W$ and WV^{-1} is an invertible linear map from X to Y. If, conversely, $A \in L(X, Y)$ is invertible, then so is the column map AV (as a map to Y), hence $\dim Y = \#AV = \#V = \dim X$. □

Discussion

Some (e.g., Hans Schneider) would prefer not to use elimination, particularly in the proof of (4)Lemma. It is certainly possible to prove (4)Lemma by first proving (17), then use an inductive argument (see, e.g., [SB; proof of (3.4.4) Theorem]) to prove (4) from the consequence of (17) that if U is 1-1 and $[U, W]$ is onto but not 1-1, then one may construct W_1 by dropping some column(s) from W and still have $[U, W_1]$ onto. I prefer not to go this route because it is lengthy, its details are very close to doing elimination, yet I would be missing the opportunity to impress upon the students the fundamental character of (1)Theorem.

The above definition of 'basis' for the vector space Y as any invertible column map to Y fails to cover infinite-dimensional vector spaces. This was done above only for the sake of simplicity. By enlarging the notion of 'coordinate space' in the customary way to include also the spaces $\mathbb{F}_0^T$ of all finitely supported scalar-valued maps on some arbitrary set T (with pointwise addition and multiplication by a scalar), the general case is covered. In fact, it is often helpful to admit such coordinate spaces even with a finite T since, in many practical situations, there is often no natural way to order a given basis. The space of all polynomials of degree $\le k$ in several variables provides a ready illustration.

It is possible to stress the ideas of duality by studying in just as much detail the second factor in a factorization $A = V\Lambda$. Such a map $\Lambda \in L(X, \mathbb{F}^n)$ is characterized by an n-sequence of linear functionals on X. If also X is a coordinate space, this corresponds to looking at the *matrix* Λ in terms of its rows rather than its columns.

Finally, the heavy use of factorizations in the above development gives the student an early taste of what is to come, as, in one view, the task of an Applied Linear Algebra course is to teach the student the use of (matrix) factorizations (such as LU, QR, SVD, similarity, and congruence).

References

[SB] H. Schneider and G.P. Barker, *Matrices and Linear Algebra*, 2nd ed., Dover Reprint, 1989.