Finite Dimensional Inner Product Spaces

Jason Kenyon

November 2022

Contents

Preface

Hello

Chapter 1

Vector Spaces

1.1 test

Theorem 1.1.1. *test*

Chapter 2

Linear Functions

Chapter 3

Linear Systems of Equations

3.1 Rank

Definition 3.1.1. An elementary row or column operation on an $m \times n$ matrix A is defined as one of the following:

1. Interchanging any two rows or columns of A
2. Scaling each entry in a row or or column of A
3. Adding a multiple of one row or column to another row or column of A

An elementary matrix is the result of applying one of the above to the $n \times n$ identity matrix.

Theorem 3.1.1. *Suppose that B is the result of applying an elementary row operation to A. Then there exists an elementary matrix E such that $B = EA$. Furthermore, E is the matrix obtained by performing the same elementary row operation to I_n as was performed to convert A into B. Similarly, if B is the result of applying an elementary column operation to A, then there exits an elementary matrix E such that $B = AE$, and E is the result of applying the same elementary column operation to I_m as was applied to A.*

The proof is a tedious verification of cases; the elementary matrices are defined precisely for this to work.

Definition 3.1.2. The rank of a matrix A is defined as the rank of the linear function $L_A = Ax$

Theorem 3.1.2. *Let $T : V \to W$ be linear and $A = [T]_\beta^\gamma$. Then* $\operatorname{rank}(T) = \operatorname{rank}(L_A)$

Proof. Consider the map $\phi_\beta : V \to \mathbb{F}^n$. That is, the function mapping a vector to its representation in coordinates. This is linear by definition and invertible as we know that any basis represents a vector uniquely as a linear combination of its elements. We have

$$L_A(\mathbb{F}^n) = L_A\phi_\beta(V) = \phi_\gamma(T(V)).$$

It follows that

$$\dim(\operatorname{im}(L_A)) = \dim(\operatorname{im}(T))$$

because ϕ_γ is an isomorphism. ∎

Theorem 3.1.3. *Let A be an $m \times n$. Let P and Q be invertible $m \times m$ and $n \times n$ matrices, respectively. Then*

1. $\operatorname{rank}(AQ) = \operatorname{rank}(A)$
2. $\operatorname{rank}(PA) = \operatorname{rank}(A)$
3. $\operatorname{rank}(PAQ)$

Proof.

$$\begin{aligned} \operatorname{im}(L_{AQ}) &= \operatorname{im}(L_A L_Q) && (3.1)\\ &= L_A L_Q(\mathbb{F}^n) && (3.2)\\ &= L_A(L_Q((\mathbb{F}^n)) && (3.3)\\ &= L_A(\mathbb{F}^n) && (3.4)\\ &= \operatorname{im}(L_A) && (3.5) \end{aligned}$$

Thus, $\operatorname{rank}(L_{AQ}) = \operatorname{rank}(L_A)$. Similarly, $\operatorname{im}(L_P L_A) = L_P(\operatorname{im}(L_A)) = \operatorname{im}(L_A)$ and so $\dim(\operatorname{im}(L_P L_A)) = \dim(\operatorname{im}(L_A))$ since P is an isomorphism. It follows, by applying the previous two results that $\operatorname{rank}(PAQ) = \operatorname{rank}(A)$. ∎

Theorem 3.1.4. *Let*

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{mn} \end{pmatrix}.$$

Then $\operatorname{rank}(A) = \dim\left(\operatorname{span}\left\{\begin{pmatrix} a_{11} \\ \vdots \\ a_{1m} \end{pmatrix}, \cdots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\}\right)$

Proof.

$$
\begin{aligned}
\operatorname{im}(L_A) &= L_A(\mathbb{F}^n) && (3.6)\\
&= L_A(\operatorname{span}\{e_1, \dots e_n\}) && (3.7)\\
&= \operatorname{span}\{Ae_1, \dots, Ae_n\} && (3.8)\\
&= \operatorname{span}\left\{\begin{pmatrix} a_{11} \\ \vdots \\ a_{1m} \end{pmatrix}, \cdots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\right\} && (3.9)
\end{aligned}
$$

Furthermore, $\dim(\operatorname{span}(X))$ is nothing but the number of linearly independent vectors in X for any set of vectors X. Thus we have shown that the rank of a matrix is nothing but the number of linearly independent vectors in its columns. ∎

Theorem 3.1.5. *Let A be an $m \times n$ matrix. Then a finite composition of elementary row and column operations applied to A results in a matrix of the form*

$$\begin{pmatrix} I_{\operatorname{rank}(A)} & O_1 \\ O_2 & O_3 \end{pmatrix}$$

where O_1, O_2, O_3 are zero matrices.

Proof. First, note that if A is a zero matrix, then by theorem 3.1.4 $\operatorname{rank}(A) = 0$, and so $A = I_0$, the degenerate case of our claim. Suppose otherwise. We proceed by induction on m, the number of rows of A. In the case that $m = 1$, we may convert A to a matrix of the form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}$$

by first making the leftmost entry 1 and adding the corresponding additive inverses of the others to the other columns. Clearly the rank of the above matrix is 1 and is of the form

$$\begin{pmatrix} I_1 & O \end{pmatrix}$$

This is another degenerate case, as it lacks zeros below the identity. Now suppose that our theorem holds when A has $m - 1$ rows.

To demonstrate that our theorem holds when A is an $m \times n$ matrix, notice that when $n = 1$, we can argue that our theorem holds as before, but using row operations instead of column operations. This is another degenerate case. For $n > 0$, note that there exists an entry $A_{ij} \neq 0$ and by applying at most an elementary row and column operation, we can move A_{ij} to position $1, 1$. Additionally, we may transform A_{ij} to value 1, and as before, transform all of the entries in row and column 1 besides A_{ij} to 0. Thus we have a matrix of the form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & x_{11} & \cdots & x_{1\ n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_{m-1\ 1} & \cdots & x_{m-1\ n-1} \end{pmatrix}$$

■

The submatrix defined by x_{ij} is of dimension $m - 1 \times n - 1$ and so must have rank $\text{rank}(A) - 1$ as elementary opertations preserve rank and deleting a row and column of a matrix reduces its rank by 1. Furthermore, by our induction hypothesis the above matrix may be converted via a finite number of elementary operations to a matrix of the form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & I_{\text{rank}(A)-1} & O_1 & \\ \vdots & & & \\ 0 & O_2 & O_3 & \end{pmatrix}$$

Therefore, for an $m \times n$ matrix A, a finite number of elementary operations converts it into a matrix of the form

$$\begin{pmatrix} I_{\text{rank}(A)} & O_1 \\ O_2 & O_3 \end{pmatrix}$$

.

Theorem 3.1.6. *For any matrix A,* $\text{rank}(A^T) = \text{rank}(A)$.

Proof. By theorem 3.1.5, we may convert A to a matrix $D = BAC$ where $B = E_1 \cdots E_p$ and $C = G_1 \cdots G_q$ where E_i and G_i are elementary row and column matrices respectively. It follows that $D^T = C^T A^T B^T$, whence

$\mathrm{rank}(A^T) = \mathrm{rank}(D^T)$ by theorem (insert) because elementary matrices are invertible, and so is the transpose of the compositions thereof. Further, D^T must be of the same form as D since the only nonzero entries of D are along the diagonal from entry $1,1$ to entry $\mathrm{rank}(A), \mathrm{rank}(A)$. Hence, we have $\mathrm{rank}(A)$ linearly independent columns in the matrix D^T.

Since the columns of D^T are the rows of D, we see that the number of linearly independent columns of A is equal to the number of linearly independent columns of A^T. In other words, the dimension of the space generated by the columns of A is equal to the dimension of the space generated by its rows. ■

Theorem 3.1.7. *Let A be an invertible $n \times n$ matrix. Then A is a product of elementary matrices.*

Proof. By the dimension theorem, if A is invertible, then $\mathrm{rank}(A) = n$. So by theorem 3.1.5 A may converted into a matrix of the form $I_n = E_1 \cdots E_p A G_1 \cdots G_q$, whence $A = E_1^{-1} \cdots E_p^{-1} I_n G_1^{-1} \cdots G_q^{-1}$. ■

Theorem 3.1.8. *Let $T : V \to W$ and $U : W \to Z$. Then*

1. $\mathrm{rank}(TU) \leq \mathrm{rank}(U)$
2. $\mathrm{rank}(TU) \leq \mathrm{rank}(T)$

Proof. We have

$$\begin{aligned} \mathrm{rank}(TU) &= \dim(\mathrm{im}(TU)) && (3.10)\\ &= \dim(\mathrm{im}(T(U(V)))) && (3.11)\\ &\subseteq U(W) && (3.12)\\ &= \mathrm{im}(U) && (3.13) \end{aligned}$$

Therefore, $\dim(\mathrm{im}(TU)) \leq \dim(\mathrm{im}(U))$. Next, let β, γ, ϕ be ordered bases for V, W, and Z, respectively; and let $A = [T]_\beta^\gamma$ and $B = [U]_\gamma^\phi$. By theorem 3.1.6

$$\begin{aligned} \dim(\mathrm{im}(TU)) &= \dim(\mathrm{im}(AB)) && (3.14)\\ &= \dim(\mathrm{im}((AB)^T)) && (3.15)\\ &= \dim(\mathrm{im}(B^T A^T)) && (3.16)\\ &\leq \dim(\mathrm{im}(A^T)) && (3.17)\\ &= \dim(\mathrm{im}(A)) && (3.18)\\ &= \dim(\mathrm{im}(T)) && (3.19) \end{aligned}$$

■

3.2 Form

We now apply the fruits of our investigation into vector spaces and linearity to solve systems of linear equations.

Definition 3.2.1. A linear system of equations is a collection of m equations of the form:

$$a_1x_1 + \cdots + a_nx_n = b$$

where $a_i, x_i, b \in \mathbb{F}$ for $1 \leq i \leq n$. Equivalently, we may say $Ax = b$ for an $m \times n$ matrix A, where $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$. If $b = \mathbf{0}$, the linear system is said to be homogenous.

Definition 3.2.2. A solution to a linear system is a vector $s \in \mathbb{F}^n$ such that $As = b$

Theorem 3.2.1. *Let A be an $m \times n$ matrix over $\mathbb{F}$. If $m < n$, then the homogenous system $Ax = 0$ has a nontrivial solution.*

Proof. Notice that, the solution set to the system $Ax = 0$ is $\ker(L_A)$, so by the dimension theorem, $\dim(\ker(A)) = n - \text{rank}(L_A)$. Additionally, we know that $\text{rank}(A)$ is nothing but the number of linearly independent vectors defined by its rows which certainly cannot exceed m. Therefore $\text{rank}(A) \leq m < n$, in which case $n - \text{rank}(A) = \dim(\ker(A)) > 0$, and so $\ker(A) \neq \{0\}$. ■

Theorem 3.2.2. *For any solution s to the linear system $Ax = b$,*

$$\{s + s_0 : As_0 = \mathbf{0}\}$$

is its solution set.

Proof. Suppose that $As = b$ and $As' = b$. Then $A(s' - s) = As' - As = b - b = 0$. It follows that $s + (s' - s) \in S$. Conversely, if $y \in S$, then $y = s + s'$, in which case $Ay = A(s + s') = As + As' = b + 0 = b$. That is, $Ay = b$. ■

Theorem 3.2.3. *Let $Ax = b$ for an $n \times n$ matrix A. If A is invertible, then the system has a single solution $A^{-1}b$. If the system has a single solution, then A is invertible.*

Proof. Suppose A is invertible. Then $A(A^{-1}b) = AA^{-1}(b) = b$. Furthermore, if $As = b$ for some $s \in \mathbb{F}^n$, then $A^{-1}(As) = A^{-1}b$ and so $s = A^{-1}b$. Next, suppose that the system has a unique solution s. Then by theorem 3.2.2, we know that the solution set $S = \{s + s_0 : As_0 = 0\}$. But this is only the case if $\ker(A) = \{0\}$, lest s not be unique. And so, by the dimension theorem, A is invertible. ∎

Theorem 3.2.4. *The linear system $Ax = b$ has a nonempty solution set if and only if* $\text{rank}(A) = \text{rank}(A|b)$.

Proof. If the system has a solution, then $b \in \text{im}(L_A)$. Additionally, $\text{im}(L_A) = L_A(F^n)$ and $L_A(e_i) = Ae_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix}$. Therefore, since $L_A(\mathbb{F}^n) = \text{span}\{Ae_1, \dots Ae_n\}$, $\text{im}(L_A) = \text{span}\{A_1, \dots A_n\}$, where A_i is the i^{th} column of A. Certainly, $b \in \text{span}\{A_1, \dots A_n\}$ if and only if $\text{span}\{A_1, \dots A_n\} = \text{span}\{A_1, \dots A_n, b\}$, which is to say $\dim(\text{im}(\text{span}\{A_1, \dots A_n\})) = \dim(\text{im}(\text{span}\{A_1, \dots A_n, b\}))$, or, $\text{rank}(A) = \text{rank}(A|b)$. ∎

Corollary 3.2.1. *Let $Ax = b$ be a linear system of m equations in n variables. Then its solution set is either, empty, of one element, or of infinitely many elements (provided that $\mathbb{F}$ is not a finite field).*

Proof. By theorem 3.2.4 $Ax = b$ has a nonempty solution set if and only if $\text{rank}(A) = \text{rank}(A|b)$. Therefore, it may be that our linear system has no solutions; however, supposing that this is not the case, by theorem 3.2.3 it has a unique solution if and only if A is invertible. Finally, assume that our linear system has neither no solution nor a single solution. This yields

$$Ax_1 = Ax_2 = b \tag{3.20}$$

for $x_1, x_2 \in \mathbb{F}^n$, which implies

$$\begin{aligned} Ax_1 - Ax_2 &= \mathbf{0} && (3.21) \\ &= A(x_1 - x_2) && (3.22) \\ &= nA(x_1 - x_2) && (3.23) \\ &= A(n(x_1 - x_2)) && (3.24) \\ & && (3.25) \end{aligned}$$

where $n \in \mathbb{F}$. Thus, by theorem 3.2.2

$$A(x_1 + n(x_1 - x_2)) = b.$$

■

3.3 Solution

Definition 3.3.1. A matrix of the form

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

is said to be in reduced echelon form if

1. $a_{ii} \neq 0$ implies that $a_{ij} = 1$
2. $a_{ij} \neq 1$ implies that $a_{ij} = 0$
3. $a_{ij} = 0$ for all $1 \leq j \leq n$ implies that $i < r$ for all nonzero rows $\begin{pmatrix} a_{r1} & \cdots & a_{rn} \end{pmatrix}$

Theorem 3.3.1. *Any matrix can be converted into reduced echelon form via a finite number of elementary row operations.*

Proof. This is a restatement of theorem 3.1.5. ■

This form is of particular interest because reducing an augmented matrix is equivalent to solving a linear system of equations. We now have a procedure for solving arbitrary systems of linear equations. For example, we may now demonstrate that a set of vectors is linearly dependent by finding a nontrivial solution to a linear system of equations; similarly we may apply theorem 3.2.4 to demonstrate that a set of vectors is linearly dependent. In the following chapter, we will also see that computing the elements of an eigenspace is made possible by reducing a matrix. It follows that

Corollary 3.3.1. *For any invertible $n \times n$ matrix A.*

$$A^{-1}(A|I_n) = E_1 \cdots E_p(A|I_n) = (I_n|A^{-1})$$

where $E_1, \ldots, E_p$ are elementary matrices.

Notice that the above elementary matrices may be either row or column matrices; however, since we are left multiplying, the product will result in a row operation. Thus we now have a procedure for finding the inverse of any matrix: perform row operations to convert it into the identity matrix, while accounting for each change. Additionally,

Corollary 3.3.2. *Let A be an $m \times n$ matrix and C be an invertible $n \times n$ matrix. Then the solutions sets to the linear systems*

$$Ax = b and CAx = Cb$$

are equal.

This follow directly from the invertibility, and fits with our intuition: as we row reduce a linear system, its solutions do not change.

Chapter 4

Eigenspaces

Chapter 5

Orthogonality

5.1 Inner Products

Hello

5.2 Projections

Definition 5.2.1. Let $V = W_1 \oplus W_2$. A projection of V on W_1 along W_2 is a linear function $T : V \to V$ such that for any $x \in V$ where $x = x_1 + x_2$ $x_1 \in W_1$ and $x_2 \in W_2$ $T(x) = x_1$.

Theorem 5.2.1. *A linear function $T : V \to V$ is a projection of V on $W_1 = \{x : T(x) = x\}$ along $\ker T$ if and only if $T = T^2$.*

Proof. If T is a projection, then clearly $T = T^2$ by definition. Conversely, for $x \in V$ we know that $x = Tx + (x - Tx)$. But by assumption $T^2x = Tx$, which means $T(Tx - x) = T(x - Tx)\mathbf{0}$. That is, $x - Tx \in \ker(T)$. Hence, $V = \{x \in V : Tx = x\} \oplus \ker(T)$ as $Tx = x$ and $Tx = 0$ implies $x = 0(x \in \ker(T))$. And so for $x \in V$, we have $x = y + z$ for $y \in \{x \in V : Tx = x\}$ and $z \in \ker(T)$, and so $Tx = Ty + Tz = y$. ■

5.3 Orthogonal Projection

Definition 5.3.1. Let $W \subseteq V$. The orthogonal complement of W is defined as $W^\perp = \{v \in V : \langle v, w \rangle = 0 \text{ for all } w \in W\}$.

Theorem 5.3.1. *The following statments are true*

1. $W^\perp$ *is a subspace of* V
2. $\dim(W^\perp) = \dim(V) - \dim(W)$

Proof. Firstly, note that $\langle \mathbf{0}, w\rangle = \mathbf{0}$ for all $w \in W$, so $\mathbf{0} \in W^\perp$. Furthermore, if $\langle w, c\rangle = 0$ for some $w \in W$ then $\langle aw, c\rangle = a\langle w, c\rangle = 0$ by linearity. Similarly, if $\langle w, a\rangle = 0$ and $\langle b, c\rangle = 0$ then $\langle w, a\rangle + \langle b, c\rangle = \langle w + b, c\rangle = 0$. Secondly, ∎

Theorem 5.3.2. *Let* $W \subseteq V$*.Then for any* $x \in V$ *there exist unique vectors* $y \in W$ *and* $z \in W^\perp$ *such that* $x = y + z$*. Furthermore, for all* $w \in W$ *s*

$$||y - x|| \leq ||w - x||$$

and we call y *the orthogonal projection of* z *on* w*, denoted* x_w*. Similarly,* z *is denoted* $x_\perp$*.*

Proof. trivial ∎

Theorem 5.3.3. *Let* $W \subseteq V$ $x \in V$ *and* $\beta = \{v_1, \dots v_n\}$ *be an orthonormal basis for* W *and* A *be the matrix whose* j^{th} *column is* v_j*. Then the orthogonal projection of* x *on* W $x_w = AA^*x$*.*

Proof. We begin by demonstrating that $W^\perp = \ker A^*$. We have

$$A^*x = \begin{pmatrix} v_1^*x \\ \vdots \\ v_n^*x \end{pmatrix} = \begin{pmatrix} \langle v_1, x\rangle \\ \vdots \\ \langle v_n, x\rangle \end{pmatrix}.$$

Certainly $Ax = \mathbf{0}$ if and only if $\langle v_i, x\rangle = 0$ for all $1 \leq i \leq n$. But that is to say $x \in W^\perp$, and so

$$\ker(A^*) = W^\perp.$$

Note that $Ax = \operatorname{span}\beta$ by definition. Therefore, for some $c \in \mathbb{F}^n$ $Ac = x_w$, which means that $x - x_W = x - Ac \in W^\perp$. It follows that $A^*(x - Ac) = 0$ and so

$$A^*Ac = A^*x.$$

Thus, if we see that $x_w = Ac$. Furthermore, since β is orthonormal, A must be unitary, in which case

$$Ac = AA^*x = x_W.$$

∎

Corollary 5.3.1. *AA^* is a projection and* $\ker(AA^*) = W^\perp$*. Additionally, AA^* is the unique such linear function.*

Proof. Surely AA^* is linear, and since we know that $x = x_W + x_{W^\perp}$ for all $x \in V$ it follows that $(AA^*)^2 x = AA^* x_W = x_w = AA^* x$. Thus the orthogonal projection is, in fact, a projection on $W^\perp = \{x \in V : AA^* x = x\}$ along $\ker(AA^*)$, by theorem 5.2.1 ($V = W \oplus W^\perp$). Furthermore, if $x = x_W + x_{W^\perp}$ with $x_W = 0$, $AA^* x = x_W = 0$. The converse follows in the same way. Thus, $\ker(AA^*) = W^\perp$ Similarly, we have $\operatorname{im}(AA^*) = W$. Additionally, as a projection is defined uniquely in terms of its range, it is clear that any other projection T on $W = \{x \in V : T(x) = x\}$ must be the same as AA^*. ■

Definition 5.3.2. There exists

Appendix A

Determinants as Permutations

Hello