1 Range Space

Definition 1. Given a linear map $L : U \to V$, we call $R(L) := \{v \in V \mid v = L(u), \ u \in U\}$ the range space of $L$. Define $\text{rank}(L) := \dim R(L)$

The range of a linear map $R(A)$ is:

- the set of vectors that can be actually reached by linear mapping $y = A(x)$
- the span of the columns of matrix $A$
- the set of vectors $y$ for which $y = A(x)$ has a solution

Fact. A linear map $A$ is called surjective (onto) if $R(A) = V$. This happens iff:

- for any $y$, $Ax = y$ can be used to solve for $x$
- rows of $A$ are independent
- $A$ has a right inverse

Recall. What determines the dimensionality of a space?

Problem 1 (Deriving range and nullspace). Let $\mathbb{R}^n[s]$ be the vector space of polynomials in $s$ whose degree is less than or equal to $n$. Consider a linear map $L : \mathbb{R}^n[s] \to \mathbb{R}^n[s]$ with $L(u) = \frac{dn}{ds}$. Specify the range and null space of $L$ and their dimensions. Is this linear map bijective?

—— group work ——

Theorem (Rank-Nullity Theorem). For a linear map $L : U \to V$, let $n$ be the dimension of $U$. $\dim(R(L)) + \dim(N(L)) = n$.

2 Invertibility

Fact. A linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ is invertible or nonsingular if $\det(A) \neq 0$. Equivalent conditions:

- columns of $A$ are a basis for $\mathbb{R}^n$, rows of $A$ are a basis for $\mathbb{R}^n$
- $y = Ax$ has a unique closed-form solution $x$ for every $y \in \mathbb{R}^n$, and $x = A^{-1}y$ is that solution
- $A$ has a (left and right) inverse denoted $A^{-1} \in \mathbb{R}^{n \times n}$, $AA^{-1} = A^{-1}A = I$
- Trivial null space: $N(A) = \theta$ due to injectivity
- $R(A) = \mathbb{R}^n$ due to surjectivity
• Others?

**Problem 2** (Matrix size, injectivity, and surjectivity). a) Can a square matrix be injective but not surjective? b) Consider \( Ax = y \) where \( A \) is "fat matrix", is the map \( A \) injective, surjective? c) What if \( A \) is a skinny matrix?

## 3 Matrix Representation

**Fact:** Any linear map over finite dimensional vector spaces can be represented by matrix multiplication.

**Basic Idea:** Let \( U, V \) be finite dimensional vector spaces and let \( A : U \to V \) be a linear map. Consider bases \( \{u_j\}_{j=1}^n \) of \( U \) and \( \{v_i\}_{i=1}^m \) of \( V \). For each \( j = 1, \ldots, n \), there exists a unique \( \{a_{1j}, \ldots, a_{mj}\} \) such that

\[
A(u_j) = \sum_{i=1}^m a_{ij}v_i.
\]

Then we can define an \( m \times n \) matrix \( A \), whose \( (i,j) \) th element is \( a_{ij} \). In other words, the \( j \) th column of \( A \) is \( A(u_j) \) expressed with respect to \( \{v_i\}_{i=1}^m \).

**Problem 3** (Matrix representation). Let \( \mathbb{R}^n[s] \) be the vector space of polynomials in \( s \) whose degree is less than or equal to \( n \). Consider a linear map \( A : \mathbb{R}^n[s] \to \mathbb{R}^{n-1}[s] \) with \( A(u) = \frac{du}{ds} \). Define bases of the domain and codomain, and give the matrix representation of \( A \).

## 4 Change of Basis

\( \{u_i\}_{i=1}^n, \{\bar{u}_i\}_{i=1}^n \) : two bases of \( U \).
\( \{v_j\}_{j=1}^m, \{\bar{v}_j\}_{j=1}^m \) : two bases of \( V \).

\( A \): matrix representation of \( A \) with \( \{u_i\}_{i=1}^n, \{v_j\}_{j=1}^m \).
\( \bar{A} \): matrix representation of \( A \) with \( \{\bar{u}_i\}_{i=1}^n, \{\bar{v}_j\}_{j=1}^m \).

\( R \) (or \( \bar{R} \)): \( n \times m \) matrix whose \( i \) th column is \( u_i \) (or \( \bar{u}_i \)).

\( S \) (or \( \bar{S} \)): \( m \times m \) matrix whose \( j \) th column is \( v_j \) (or \( \bar{v}_j \)).

For any \( x \in U \), \( x = R\xi = \bar{R}\xi \implies \xi = P\xi \), where \( P = R^{-1}\bar{R} \).

For \( y = A(x) \in V \), \( y = S\eta = \bar{S}\eta \implies \eta = Q\eta \), where \( Q = \bar{S}^{-1}S \).

In addition, \( \eta = A\xi \).
\[ \therefore \quad \bar{\eta} = QAP\xi \implies \bar{A} = QAP \]
Recall. What is the relationship between a similarity transform and change of basis?

Problem 4 (Change of basis matrix). Consider the basis $B = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$

What is the change of basis matrix from the basis $B$ to the standard basis? What about in the opposite direction?

More on change of basis: https://www.youtube.com/watch?v=P2LTAUO1TdA&t=175s&ab_channel=3Blue1Brown

5 Vector Norms

Let $(V,F)$ be a vector space with $F = \mathbb{R}$ or $F = \mathbb{C}$. A norm on that space is a map $\| \cdot \| : V \to \mathbb{R}_+$ satisfying the following axioms:

1. $\|\alpha v\| = |\alpha| \|v\|$ $\forall \alpha \in F, v \in V$ (absolute homogeneity)
2. $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$ $\forall v_1, v_2 \in V$ (subadditivity or triangle inequality)
3. $\|v\| = 0 \iff v = \theta_V$

Consider. We define the codomain of $\| \cdot \|$ as $\mathbb{R}_+$. Do the axioms above imply that $\| \cdot \|$ always produces a non-negative result?

Consider. Why are norms useful?

Two norms $\| \cdot \|_a$ and $\| \cdot \|_b$ on $(V,F)$ are said to be equivalent if $\exists m_l, m_u \in \mathbb{R}_+$ such that $\forall v \in V$:

$$m_l \|v\|_a \leq \|v\|_b \leq m_u \|v\|_a$$

Fact 1. All norms over finite dimensional vector spaces are equivalent.

Consider. Why is it helpful to have equivalent norms on a vector space?
6 Induced norms

Let $U$ and $V$ be normed linear spaces with norms $\| \cdot \|_U$ and $\| \cdot \|_V$, respectively.

**Definition 2.** The *induced norm* of a continuous linear operator $A : U \to V$ is given by

$$\|A\|_i := \sup_{u \neq \theta_U} \frac{\|A(u)\|_V}{\|u\|_U} = \sup_{\|u\|_U = 1} \|A(u)\|_V$$

**Consider.** Norms on two vector spaces $U$ and $V$ “induce” a norm on the space of linear maps between $U$ and $V$. *(Exercise: Prove that the space of linear maps between two vector spaces is a vector space.)* In fact, every pair of norms $(\| \cdot \|_U, \| \cdot \|_V)$ may induce a different norm on that space.

**Proposition 3.** Let $U, V, W$ be normed linear spaces. Given the linear operators $A, \tilde{A} : V \to W$, and $B : U \to V$, we have the following matrix norm properties:

1. $\|A(v)\|_W \leq \|A\|_i \|v\|_V$
2. $\|\alpha A\| = |\alpha| \|A\|$
3. $\|A + \tilde{A}\|_i \leq \|A\| + \|\tilde{A}\|$
4. $\|A\| = 0 \iff A = 0$
5. $\|AB\| \leq \|A\| \|B\|$

**Problem 5** (Induced norms of matrices). Given a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$, we let $A \in \mathbb{R}^{m \times n}$ be its matrix representation. Then the induced norm of $A$ is defined by

$$\|A\|_{p,i} := \sup_{u \neq 0} \frac{\|Au\|_p}{\|u\|_p}$$

Show that

$$\|A\|_{1,i} = \max_{j=1}^{m} \left\{ \sum_{i=1}^{n} |a_{ij}| \right\} \text{ (max column sum)}$$

—— group work ——
Addendum: Using Norms for Sensitivity Analysis

Consider. Suppose we have we have \( \mathbf{y} = \mathbf{C} \mathbf{x} \) with \( \mathbf{C} \) invertible. If \( \mathbf{y} \) or \( \mathbf{C} \) is perturbed, how confident are you that \( \mathbf{C} \) is still invertible? How can we analyze how sensitive the solution \( \mathbf{x} \) is to perturbations in \( \mathbf{C} \) and \( \mathbf{y} \)?

Let \( \mathbf{x}_0 \) be the nominal solution to \( \mathbf{A} \mathbf{x}_0 = \mathbf{b} \). If we perturb \( \mathbf{x}_0 \), get \( \mathbf{x} = \mathbf{x}_0 + \delta \mathbf{x} \), and our new equation is:

\[(\mathbf{A} + \delta \mathbf{A})(\mathbf{x}_0 + \delta \mathbf{x}) = (\mathbf{b} + \delta \mathbf{b})\]

Goal: relate \( ||\delta \mathbf{x}|| \) to \( ||\delta \mathbf{A}|| \) and \( ||\delta \mathbf{b}|| \)

1. \( \mathbf{A} \mathbf{x}_0 + \delta \mathbf{A} \mathbf{x}_0 + A \delta \mathbf{x} + \delta \mathbf{A} \delta \mathbf{x} = \mathbf{b} + \delta \mathbf{b} \)

Ax0, now assume \( \delta \mathbf{A} \delta \mathbf{x} \) is negligible (zero).

2. \( \delta \mathbf{A} \mathbf{x}_0 + A \delta \mathbf{x} = \delta \mathbf{b} \)

now rearrange.

3. \( \delta \mathbf{x} = \mathbf{A}^{-1}(\delta \mathbf{b} - A \delta \mathbf{x}) \)

now using matrix norm properties (1) and (3).

4. \( \frac{||\delta \mathbf{x}||}{||\mathbf{x}||} \leq ||\mathbf{A}^{-1}|| \cdot ||\delta \mathbf{A}|| \cdot ||\mathbf{x}|| + ||\mathbf{A}^{-1}|| \cdot ||\delta \mathbf{b}|| \)

now divide by \( ||\mathbf{x}|| \)

5. \( \frac{||\delta \mathbf{x}||}{||\mathbf{x}||} \leq ||\mathbf{A}^{-1}|| \cdot ||\delta \mathbf{A}|| + \frac{||\mathbf{A}^{-1}|| \cdot ||\delta \mathbf{b}||}{||\mathbf{x}||} \)

6. We’d like this form: \( \frac{||\delta \mathbf{x}||}{||\mathbf{x}||} \leq \left[ \left( \frac{||\delta \mathbf{A}||}{||\mathbf{A}||} + \frac{||\delta \mathbf{b}||}{||\mathbf{b}||} \right) \right] \cdot ||\mathbf{A}^{-1}|| \cdot ||\mathbf{A}|| \)

so that we can quantify how changes in \( \mathbf{A} \) and \( \mathbf{b} \) affect changes in \( \mathbf{x} \)

7. Note that (1): \( \mathbf{b} = \mathbf{A} \mathbf{x} \implies ||\mathbf{b}|| \leq ||\mathbf{A}|| \cdot ||\mathbf{x}|| \implies \frac{1}{||\mathbf{x}||} \leq \frac{||\mathbf{A}||}{||\mathbf{b}||} \)

8. Thus multiply and divide by \( \frac{||\mathbf{A}||}{||\mathbf{x}||} \) and apply (1): \( \frac{||\delta \mathbf{x}||}{||\mathbf{x}||} \leq ||\mathbf{A}^{-1}|| \cdot ||\delta \mathbf{A}|| \frac{||\mathbf{A}||}{||\mathbf{A}||} + \frac{||\mathbf{A}^{-1}|| \cdot ||\delta \mathbf{b}||}{||\mathbf{b}||} \cdot ||\mathbf{A}|| \)

now factoring.

9. \( \frac{||\delta \mathbf{x}||}{||\mathbf{x}||} \leq ||\mathbf{A}^{-1}|| \cdot ||\mathbf{A}|| \left( \frac{||\delta \mathbf{A}||}{||\mathbf{A}||} + \frac{||\delta \mathbf{b}||}{||\mathbf{b}||} \right) \)

The quantity \( k = ||\mathbf{A}^{-1}|| \cdot ||\mathbf{A}|| \) is called the condition number. It is important because if \( k \gg 1 \), \( \mathbf{A} \) is invertible but not very invertible (so we say \( \mathbf{A} \) is ”poorly conditioned”), and small changes in \( \mathbf{A} \) and \( \mathbf{b} \) amplify large changes in \( \mathbf{x} \). Conversely, if \( k \approx 1 \), \( \mathbf{A} \) is highly invertible (so we say \( \mathbf{A} \) is ”well conditioned”), and perturbations have little affect on the solution \( \mathbf{x} \). You may have seen MATLAB warnings before when MATLAB is trying to invert a poorly conditioned matrix because it runs into numerical difficulties.

Final note: later when we cover SVD and mention pseudoinverses, you can consider sensitivity when \( \mathbf{A} \) is not invertible. In that case, this analysis quantifies how close to invertible \( \mathbf{A} \) is, and we use the pseudoinverse \( \mathbf{A}^\dagger \) instead of \( \mathbf{A}^{-1} \) in these steps.