

Review of Basic Vector Space and Matrix Properties

- In Set 1 of the notes we review some basic definitions and properties needed for the numerical linear algebra discussion. We also phrase them in terms of certain matrix properties and decompositions.
- Reading related to these topics
 - GV96 All of Chapter 1 and in Chapter 2: 2.1 – 2.3, 2.5, and 2.6
 - Stew98 in Chapter 1: 1.1 – 2.5, 3.1 – 3.4, 4.1 – 4.5
- The notes adopt Stewart's notational convention of using Greek lower case letters for scalars (complex or real); Upper case Roman or Greek letters for matrices; Lower case Roman for vectors. Note GV96 does not always adhere to this standard.

Vector Spaces, Matrices, and Systems of Equations

We start by assuming familiarity with the field of real numbers (\mathfrak{R}) and the field of complex numbers (\mathcal{C})

- complex number: $\alpha = \beta + i\gamma$ where i here is used to represent the root of -1 (occasionally we will use j for this but it will be made clear when this is done)
- β and γ are the real and imaginary parts of α respectively
- complex conjugate $\bar{\alpha} = \beta - i\gamma$
- the absolute value of α denoted $|\alpha|$ is $\sqrt{\alpha\bar{\alpha}} = \sqrt{\beta^2 + \gamma^2}$

Given a field of scalars \mathcal{F} , (we will assume \mathcal{F} is \mathbb{R} or \mathbb{C}) we can define a vector space.

Definition: Given scalars \mathcal{F} , a set of vectors \mathcal{V} , a vector addition operation $x + y$ for $x, y, z \in \mathcal{V}$, and a scalar-vector product operation $y = \alpha x$ for $x, y \in \mathcal{V}$ and $\alpha \in \mathcal{F}$, we have a vector space if the following properties hold:

$$\begin{aligned}x + y &= y + x & (1) \\(x + y) + z &= x + (y + z) & (2) \\x + 0 &= x & (3) \\x + (-1_s)x &= 0 & (4) \\(\alpha\beta)x &= \alpha(\beta x) & (5) \\(\alpha +_s \beta)x &= \alpha x + \beta x & (6) \\\alpha(x + y) &= \alpha x + \alpha y & (7) \\1_s x &= x & (8)\end{aligned}$$

Note that the 0 in properties 3 and 4 is a vector in \mathcal{V} and the ± 1 in properties 4 and 8 is a scalar from \mathcal{F}

Note there is also a scalar $0 \in \mathcal{F}$ that plays a role. (identify which 0s are scalars and vectors in the proof.) This can be deduced from the properties in the definition and the proof is an example of the use of these properties to deduce relationships that are consistent with what we would expect based on intuition.

$$\begin{aligned} 0 &= a + (-1)a \text{ prop4} \\ &= 1a + (-1)a \text{ prop8} \\ &= (0 + 1)a + (-1)a \text{ scalar } 0 + 1 = 1 \\ &= (0a + 1a) + (-1)a \text{ prop6} \\ &= 0a + (1a + (-1)a) \text{ prop2} \\ &= 0a + (a + (-1)a) \text{ prop8} \\ &= 0a + (0) \text{ prop4} \\ &= 0a \text{ prop3} \end{aligned}$$

Examples

- \mathfrak{R}^n – a vector is an ordered list of n real scalars
 - addition of vectors is componentwise scalar addition
 - scalar vector product multiplies each component of the vector with the scalar
 - \mathcal{C}^n defined analogously
- \mathcal{P}_n – the set of polynomials of degree less than or equal to n
 - isomorphic to \mathcal{C}^{n+1}
 - elements can be written as a linear combination of $n+1$ monomials therefore finite dimensional space
- \mathcal{P}_∞ – the set of polynomials of any degree
 - any element can be written as a finite sum of monomials
 - infinite dimensional since it is not the same finite sum size for all vectors
- $\mathcal{C}[0, 1]$ – continuous functions on the interval $[0, 1]$
 - infinite dimensional
 - need concept of convergence to discuss infinite linear combination that represents each vector

We are concerned in the first portion of the course with \mathfrak{R}^n and \mathcal{C}^n . Consider \mathfrak{R}^3 :

$$x = \begin{pmatrix} 1 \\ 3 \\ -52 \end{pmatrix}$$

$$y = \begin{pmatrix} 10 \\ -4 \\ 2 \end{pmatrix}$$

$$x + y = \begin{pmatrix} 11 \\ -1 \\ -50 \end{pmatrix}$$

$$2x = \begin{pmatrix} 2 \\ 6 \\ -104 \end{pmatrix}$$

$$0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The following are the standard basis vectors:

$$e = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

DEFINITION: An $m \times n$ matrix of scalars from \mathfrak{R} or \mathcal{C} is a two dimensional arrangement of mn scalars

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{pmatrix}$$

The set of $m \times n$ matrices with scalar elements from \mathfrak{R} is denoted $\mathfrak{R}^{m \times n}$

The set of $m \times n$ matrices with scalar elements from \mathcal{C} is denoted $\mathcal{C}^{m \times n}$

DEFINITION: The transpose of an $m \times n$ real matrix A is denoted A^T and is an $n \times m$ real matrix

$$A^T = \begin{pmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{m1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{m2} \\ \vdots & \vdots & & \vdots \\ \alpha_{1n} & \alpha_{2n} & \cdots & \alpha_{mn} \end{pmatrix}$$

DEFINITION: The hermitian transpose of an $m \times n$ complex matrix A is denoted A^H and is an $n \times m$ complex matrix

$$A^H = \begin{pmatrix} \bar{\alpha}_{11} & \bar{\alpha}_{21} & \cdots & \bar{\alpha}_{m1} \\ \bar{\alpha}_{12} & \bar{\alpha}_{22} & \cdots & \bar{\alpha}_{m2} \\ \vdots & \vdots & & \vdots \\ \bar{\alpha}_{1n} & \bar{\alpha}_{2n} & \cdots & \bar{\alpha}_{mn} \end{pmatrix}$$

We can now relate a system of linear equations, a linear combination of vectors, and a matrix vector product operation that is the starting point for much of linear algebra.

A system of m linear equations in n unknowns can be written:

$$\begin{aligned}\alpha_{11}\xi_1 + \alpha_{12}\xi_2 + \cdots + \alpha_{1n}\xi_n &= \beta_1 \\ \alpha_{21}\xi_1 + \alpha_{22}\xi_2 + \cdots + \alpha_{2n}\xi_n &= \beta_2 \\ &\vdots \\ \alpha_{m1}\xi_1 + \alpha_{m2}\xi_2 + \cdots + \alpha_{mn}\xi_n &= \beta_m\end{aligned}$$

A linear combination of n vectors from \mathfrak{R}^n or \mathcal{C}^n is:

$$a_1\xi_1 + a_2\xi_2 + \cdots + a_n\xi_n$$

where

$$a_i = \begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \\ \vdots \\ \alpha_{mi} \end{pmatrix}$$

So if we define the matrix $A \in \mathfrak{R}^{m \times n}$

$$A = (a_1 \quad a_2 \quad \cdots \quad a_n)$$

and the vector $x \in \mathfrak{R}^n$

$$x = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}$$

we can define the matrix vector product as the linear combination of the n columns of A using the n scalars in x

$$Ax = a_1\xi_1 + a_2\xi_2 + \cdots + a_n\xi_n$$

Finally, defining the vector $b \in \mathfrak{R}^m$ as

$$b = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix}$$

we can see that the system of equations

$$\begin{aligned} \alpha_{11}\xi_1 + \alpha_{12}\xi_2 + \cdots + \alpha_{1n}\xi_n &= \beta_1 \\ \alpha_{21}\xi_1 + \alpha_{22}\xi_2 + \cdots + \alpha_{2n}\xi_n &= \beta_2 \\ &\vdots \\ \alpha_{m1}\xi_1 + \alpha_{m2}\xi_2 + \cdots + \alpha_{mn}\xi_n &= \beta_m \end{aligned}$$

can be written as a matrix vector product

$$Ax = b$$

where we are given A and b and must find x to solve the system.

Defining what it means to solve the system, the properties that characterize the solution(s), and the algorithms to efficiently and accurately determine the solution(s) is the one of main tasks of numerical linear algebra.

Matrix Operations

- $B = \gamma A$ is a matrix with elements $\beta_{ij} = \gamma \alpha_{ij}$
- $C = A + B$ is a matrix with elements $\gamma_{ij} = \beta_{ij} + \alpha_{ij}$
- defined and associative for $A, B, C \in \mathfrak{R}^{m \times n}$
- $C = AB$ is a matrix with elements $\gamma_{ij} = \sum_{k=1}^{n_2} \alpha_{ik} \beta_{kj}$
- defined when $A \in \mathfrak{R}^{n_1 \times n_2}$, $B \in \mathfrak{R}^{n_2 \times n_3}$, and results in $C \in \mathfrak{R}^{n_1 \times n_3}$
- the product is: not commutative, is associative, and is distributive, i.e., $A(B + C) = AB + AC$
- Alternative definition 1. $C = AB \rightarrow c_i = Ab_i$ for $i = 1, \dots, n_3$ where $c_i = Ce_i$, $b_i = Be_i$
- Alternative definition 2. $C = AB = \sum_{k=1}^{n_2} a_k b_k^T$ where $a_k = Ae_k$, $b_k^T = e_k^T B$

Subspaces, Bases, Linear Independence and Linear Transformations

The algebraic structure of the vector spaces \mathfrak{R}^n and \mathcal{C}^n is **common to all finite dimensional vector spaces**. We will use \mathfrak{R}^n in most of our discussions but the results can be adapted to \mathcal{C}^n and all other such vector spaces.

By definition a vector space \mathcal{V} is closed under linear combinations, but an arbitrary subset of the space is not necessarily closed.

DEFINITION: A subset $\mathcal{S} \subseteq \mathfrak{R}^n$ is a **subspace** if it is closed under linear combination, i.e., if $x_1, x_2, \dots, x_k \in \mathcal{S}$ then for any scalars $\alpha_i, i = 1, \dots, k$

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k \in \mathcal{S}$$

and in fact the subspace is itself a vector space (and hence all of our results apply within \mathcal{S}).

DEFINITION: Let $\mathcal{S} \subseteq \mathfrak{R}^n$ be a subset. The set of all linear combinations of vectors in \mathcal{S} is call the **span** of \mathcal{S} and is a subspace.

EXAMPLE: $\mathfrak{R}^n = \text{span}(e_1, e_2, \dots, e_n)$

DEFINITION: The set of vectors x_1, \dots, x_k are **linearly independent** if

$$\alpha_1 x_1 + \dots + \alpha_k x_k = 0 \rightarrow \alpha_i = 0$$

for $i = 1, \dots, k$. If this does not hold then the vectors are linearly dependent.

NOTE: A set of vectors being linearly dependent implies one of the vectors can be written as a linear combination of the others.

NOTE: Any set that contains the 0 vector is linearly dependent.

EXAMPLE:

$$x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad y = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

are linearly independent in \mathbb{R}^3 .

$$x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad y = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad z = \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix}$$

are linearly dependent

NOTE: One can often use the nonzero structure of the vectors to deduce linear independence.

DEFINITION: A set of vectors $x_1, x_2, \dots, x_k \in \mathcal{S} \subseteq \mathbb{R}^n$ is a **basis** for the subspace \mathcal{S} if

- x_1, x_2, \dots, x_k are linearly independent,
- $\text{span}(x_1, x_2, \dots, x_k) = \mathcal{S}$

LEMMA: We have the following:

- A subspace may have many bases but every basis contains k vectors and the unique integer k is the dimension of the subspace ($k = \dim(\mathcal{S})$).
- Every subspace has a basis.
- If $b \in \mathcal{S}$ and $b = Ax$ where the columns of A are a basis for \mathcal{S} then x is unique.
- Any collection of vectors in \mathcal{S} with $k + 1$ or more vectors is linearly dependent.

Matrices can also be related to bases of subspaces and linear transformations between vector (sub)spaces

Let $A \in \mathcal{C}^{m \times n}$ be an $m \times n$ matrix.

To any $x \in \mathcal{C}^n$ we can apply A via a matrix vector product and compute $b \in \mathcal{C}^m$

If we evaluate b for all possible x vectors we have in fact evaluated the span of the columns of A , i.e., all possible linear combinations of the columns of A

DEFINITION: Given $A \in \mathcal{C}^{m \times n}$

- The span of the columns of A is a subspace of \mathcal{C}^m called the **range** of A and is denoted $\mathcal{R}(A)$.
- The dimension of $\mathcal{R}(A)$ is also called the **column rank** of A .
- If the columns of A are linearly independent then they form a basis for $\mathcal{R}(A)$ and the matrix A is said to be of **full column rank**.
- Since $A(x + y) = Ax + Ay$, A defines a linear function that relates b and x and maps \mathcal{C}^n to $\mathcal{R}(A) \subseteq \mathcal{C}^m$.

The domain of the transformation defined by $A \in \mathcal{C}^{m \times n}$ is \mathcal{C}^n but, not all of this space affects the result b .

DEFINITION: The set of all vectors $x \in \mathcal{C}^n$ such that

$$Ax = 0$$

i.e., that map to the 0 vector in \mathcal{C}^m , is a subspace of \mathcal{C}^n called the **null space** of A and is denoted $\mathcal{N}(A)$.

We can therefore characterize the situation when more than one vector in \mathcal{C}^n can be mapped to the same $b \in \mathcal{C}^m$.

LEMMA: If $Ax = b$ then for any $y \in \mathcal{N}(A)$.

$$A(x + y) = b$$

We now have many of the basic questions dealt with in linear algebra.

Let $A \in \mathcal{C}^{m \times n}$, $x \in \mathcal{C}^n$ and $b \in \mathcal{C}^m$,

- When does x exist given A and b ?
- If x exists is it unique?
- If x is not unique how is the solution chosen (regularization)?
- If x does not exist can the definition of “to solve” be changed to yield something useful (approximation)?
- In all cases where a desired x exists how can it be computed efficiently and accurately?
- The term “accurately” involves both the effect of finite precision and the effect of an infinite number of computations

Norms, Distance, and Angles

In addition to the algebraic properties discussed so far we can also define analytic properties of vector spaces and the associated linear transformations.

DEFINITION: A vector norm, $\|x\|$, is a function $\mathcal{C}^n \rightarrow \Re$ that satisfies

- $x \neq 0 \rightarrow \|x\| > 0$ (definiteness)
- $\|\alpha x\| = |\alpha| \|x\|$ (homogeneity)
- $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)

We can also deduce

$$\|x - y\| \geq \left| \|x\| - \|y\| \right|$$

Examples of vector norms

Let $x \in \mathcal{C}^n$ with elements $e_i^H x = \xi_i$.

$$\begin{aligned}\|x\|_1 &= \sum_{i=1}^n |\xi_i| \\ \|x\|_\infty &= \max_{1 \leq i \leq n} |\xi_i| \\ \|x\|_2 &= \sqrt{\sum_{i=1}^n |\xi_i|^2}\end{aligned}$$

THEOREM: Let $\mu(x)$ and $\nu(x)$ be vector norms then there exist $\sigma > 0$ and $\tau > 0$ such that

$$\sigma\mu(x) \leq \nu(x) \leq \tau\mu(x)$$

In other words, for analytical purposes, all norms are equivalent.

DEFINITION: A matrix norm on $\mathcal{C}^{m \times n}$ denoted $\|A\|$ maps $\mathcal{C}^{m \times n} \rightarrow \mathfrak{R}$ and satisfies

- $A \neq 0 \rightarrow \|A\| > 0$
- $\|\alpha A\| = |\alpha| \|A\|$
- $\|A + B\| \leq \|A\| + \|B\|$

DEFINITION: The matrix norm $\|A\|$ is **consistent** if

$$\|AB\| \leq \|A\| \|B\|$$

whenever the product exists.

NOTE: A special case of consistency is

$$\|Ax\| \leq \|A\| \|x\|$$

which says that we can associate the norm of the matrix with a bound on the expansion or contraction of the linear transformation defined by A .

Examples of matrix norms

Let $A \in \mathcal{C}^{m \times n}$ with elements $e_i^H A e_j = \alpha_{ij}$.

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |\alpha_{ij}| = \max_{1 \leq j \leq n} \|A e_j\|_1$$

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |\alpha_{ij}| = \max_{1 \leq i \leq m} \|e_i^H A\|_1$$

$$\|A\|_2 = \max_{\|x\|=1} \|Ax\|_2$$

$$\|A\|_F = \sqrt{\sum_{i=1, m; j=1, n} |\alpha_{ij}|^2}$$

NOTE: The Frobenius norm $\|A\|_F$ is essentially the vector 2 norm applied to the matrix as if it was a element of \mathcal{C}^{mn} . We also have $\|A\|_F^2 = \text{trace}(A^H A)$ where the trace is the sum of the diagonal elements.

NOTE: While all matrix norms are equivalent for analytical purposes, they **differ considerably in their ease of computation.**

We can define a concept of angle between vectors that is consistent with the intuitive notion we use in \mathbb{R}^2 and \mathbb{R}^3

DEFINITION: Let x and y be two nonzero vectors in \mathcal{C}^n then the cosine of the angle between the vectors, $0 \leq \theta \leq \pi/2$, is determined by the relation

$$|x^H y| = \cos\theta \|x\|_2 \|y\|_2$$

NOTE: The cosine makes the Cauchy inequality an equality

$$|x^H y| \leq \|x\|_2 \|y\|_2$$

DEFINITION: The vectors x and y are said to be orthogonal if $x^H y = 0$.

NOTE: This generalizes the Pythagorean Theorem to multidimensional and complex vectors:

$$\begin{aligned}\|x + y\|_2^2 &= (x + y)^H (x + y) \\ &= x^H x + y^H y + 2\operatorname{Re}(x^H y) \\ &= x^H x + y^H y \\ &= \|x\|_2^2 + \|y\|_2^2\end{aligned}$$

DEFINITION: A matrix in $\mathcal{C}^{m \times m}$ (or in $\mathcal{R}^{m \times m}$) with columns $Ae_i = a_i$ is said to be **unitary** (**orthogonal**) if

- $\|a_i\|_2 = 1$ for all $i = 1, \dots, m$
- $a_i^H a_j = 0$ for $i \neq j$
- Equivalently, $AA^H = A^H A = I_m$

These matrices play an extremely important role as a accurate computational primitive and as a powerful analytical tool

If $m \neq n$ matrices that have orthonormal columns are called isometries.

DEFINITION: Let $x \in \mathcal{C}^n$ and $x_0 \in \mathcal{C}^n$ be such that

$$x^H x_0 = \gamma$$

- The set, $\mathcal{H}(x)$, of vectors y such that $x^H y = 0$ is a **hyperplane** that passes through the origin.
- $\mathcal{H}(x)$ is also a subspace of dimension $n - 1$
- The hyperplane generalizes the concept of a line in \mathbb{R}^2 and a plane in \mathbb{R}^3 .
- The hyperplane, $\mathcal{V}(x, x_0)$, of vectors y such that $x^H y = \gamma$ is a translation of a subspace.
- Note $x_0 \in \mathcal{V}(x, x_0)$ and $\mathcal{V}(x, x_0) = x_0 + \mathcal{H}(x)$.
- x_0 is not unique and usually $x_0 = \alpha x$ is used as the representative.

This yields a geometric interpretation of the solution of a system of linear equations

Consider $A \in \mathcal{C}^{m \times n}$ with rows $e_i^H A = a_i^H$, and $b \in \mathcal{C}^m$, then each row of the matrix equation $Ax = b$ defines an hyperplane i.e.,

$$\begin{aligned} a_1^H x &= \beta_1 \\ a_2^H x &= \beta_2 \\ &\vdots \\ a_m^H x &= \beta_m \end{aligned}$$

and the solution x lies in the intersection of the hyperplanes.

Consider an example in \mathfrak{R}^2

$$\begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

So we have

$$\begin{aligned} a_1^T &= (-1, 1) & \beta_1 &= 0 & \xi_2 &= \xi_1 \\ a_2^T &= (0, 1) & \beta_2 &= 2 & \xi_2 &= 2 \end{aligned}$$

The intersection is

$$x = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

which can be seen in the following picture

