Note Summary: Linear Algebra

Guoxuan*

August 13, 2022

1 Vectors and Vector Spaces

1.1 Vector Spaces

Definition 1. A Triple $(V, +, \cdot)$ where
set V, whose element are called vectors.
an operation $+ : V^2 \to V$ called vector addition (if u and v are vectors, we note $\mathbf{u} + \mathbf{v}$).
an operation $\cdot : \mathbb{R} \times V \to V$ called scalar multiplication (if $\mathbf{v}$ is a vector and λ a real, we note $\lambda\mathbf{v}$).
is said to be a (real) vector space (or a vector space over R) iff it satisfies the following 7 axioms

1. Vector addition is commutative and associative: $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V, \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$, and $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.

2. Existence of an identity element of addition: there exists an **zero vector** noted 0 s.t. $\mathbf{0} + \mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{u}$, $\forall u \in \mathbf{V}$.

3. Existence of an inverse element of addition: for any $\mathbf{u} \in \mathbf{V}$, there exists an **additive inverse** of $\mathbf{u}$, noted $-\mathbf{u}$, s.t. $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

4. Existence of an identity element of scalar multiplication: 1, a multiplicative identity in $\mathbb{R}$, s.t. $1\mathbf{u} = \mathbf{u}$.

5. Mixed associativity: for any $\lambda_1, \lambda_2 \in \mathbb{R}$ and $\mathbf{v} \in V$, we have $(\lambda_1\lambda_2)\mathbf{v} = \lambda_1(\lambda_2\mathbf{v})$.

6. Scalar multiplication is distributive w.r.t. vector addition: $\lambda(\mathbf{v_1} + \mathbf{v_2}) = \lambda\mathbf{v_1} + \lambda\mathbf{v_2}$.

7. Scalar multiplication is distributive w.r.t. addition in $\mathbb{R}$: $(\lambda_1 + \lambda_2)\mathbf{v} = \lambda_1\mathbf{v} + \lambda_2\mathbf{v}$.

*This note mainly borrows from math camp material in Columbia University, https://www.sites.google.com/site/mathcamp2018cu, and all kinds of online material. All errors are mine.

Notice that the two operators $+ : V \times V \to V$ and $\cdot : \mathbb{R} \times V \to V$ for the vector space are different from $+$ and $\cdot$ defined on $\mathbb{R}$. Also, $\mathbf{0}$ in (2) is the **zero vector**, not the neutral element 0 in $\mathbb{R}$, although we usually (ab)use the same notation.
We can show the following results using the 7 axioms of a vector space:

1. the zero vector $\mathbf{0}$ is unique in a vector space;
2. the additive inverse of vector $\mathbf{v} \in \mathbf{V}$ is unique;
3. $-\mathrm{v} = (-1)\mathrm{v}$. And therefore we can define vector subtraction by $\mathbf{v_1} - \mathbf{v_2} \coloneqq \mathbf{v_1} + (-\mathbf{v_2})$;
4. $0\mathbf{v} = \mathbf{0}$, $\forall \mathbf{v}$; $\lambda\mathbf{0} = \mathbf{0}$, $\forall \lambda \in \mathbb{R}$, and that $\lambda\mathbf{v} = 0$ implies either $\lambda = 0$ or $\mathbf{v} = \mathbf{0}$. These are left as exercises.

It is also possible to define a vector space over $\mathbb{C}$ instead of $\mathbb{R}$, in which case the definition is the same replacing $\mathbb{R}$ by $\mathbb{C}$ and we have a complex vector space. We will do so in some specific situations; unless otherwise mentioned, a vector space will be understood as a real vector space .
A major example of a n-dimensional real vector space is $\langle\mathbb{R}^n, +, \cdot\rangle$i, where the vector addition and scalar multiplication are defined in a component-by-component fashion. An element $\mathbf{v}$ in $\mathbb{R}^n$ is $\mathbf{v} = (v_1, v_2, ..., v_n)$, where $v_1 \in \mathbb{R}$ are called components or coordinates of $\mathbf{v}$. The zero vector is the vector whose components are all zero.
But the concept of vector spaces can be much more general than that. For example, the set of functions from X to R is a vector space, where X is some arbitrary nonempty set. The vector addition is defined by $(f + g)(x) \coloneqq (x) + g(x)$, and the scalar multiplication is defined by $(\lambda f)(x) \coloneqq f(x)+g(x)$. The zero vector is the function that is constant at 0. We can verify that this satisfies all the requirements for a vector space. For a vector space of $(V, +, \cdot)$, a subset W is called a vector subspace of $(V, +, \cdot)$ iff $(W, +|_W, \cdot|_W)$ is a vector space, where $+|_W$ and $\cdot|_W$ are $+$ and $\cdot$ defined for V restricted in W.

Definition 2. $(W, +, \cdot)$ is a vector subspace of $(V, +, \cdot)$ iff:

1. W contains the zero vector $\mathbf{0}$.
2. $\lambda\mathbf{u} + \mu\mathbf{v} \in W,\ \forall\ \lambda, \mu \in \mathbb{R}, \forall \mathbf{u}, \mathbf{v} \in V$

It can be shown that intersection of vector subspaces is still a vector subspace, but union may not.

Definition 3. Let $(V, +, \cdot)$ be a vector space and S a subset of V . The **linear span** of S, noted $Span(S)$, is the smallest vector subspace that contains S, that is the intersection of all the subspaces that contain S:

$$Span(S) = \cap\{W, W \text{ a vector subspace of V, and } S \subseteq W\}$$

Let $v_1, ..., v_n$ be n vectors of V. A **linear combination** of $\mathbf{v_1}, ..., \mathbf{v_n}$ is a vector $\lambda_1\mathbf{v_1} + ... + \lambda_n\mathbf{v_n}$ for n scalars $\lambda_1, ..., \lambda_n \in \mathbb{R}$.
It is straightforward that for a finite subset S if V , Span(S) is simply the set of all linear combinations of vectors in S.

1.2 Linear Dependence

Definition 4. In vector space $(V, +, \cdot)$, a finite set of vectors $\mathbf{v}_1, ..., \mathbf{v}_n$ are said to be linearly independent, iff the linear combination $\sum_{i=1}^{n} \lambda_i \mathbf{v}_i = \mathbf{0}$ implies $\lambda_i = \mathbf{0}$ for any i. Otherwise, $\mathbf{v}_1, ..., \mathbf{v}_n$ are said to be linearly dependent.

Clearly, if a vector can be represented by a linear combination of a set of linearly independent vector, then the representation is unique.

Proposition 1. *Let $\mathbf{v}_1, ..., \mathbf{v}_n$ be linearly independent elements of a vector space $(V, +, \cdot)$. Let $(\lambda_i)_i$ and $(\mu_i)_i$ be reals. If:*

$$\lambda_1 \mathbf{v_1} + ... + \lambda_n \mathbf{v_n} = \mu_1 \mathbf{v_1} + ... + \mu_n \mathbf{v_n}$$

then for all $i, \lambda_i = \mu_i$.

Proposition 2. *In a vector space, $\mathbf{v_1}, ..., \mathbf{v_n}$ are linearly dependent iff $\exists v_i, (i \in \{1, 2, ..., n\})$ that can be represented by a linear combination of $(v_j)_{j \neq i}$.*

Theorem 1. *In vector space $(V, +, \cdot)$, if $u_1, ..., u_m$ can be represented by linear combinations of $\mathbf{v_1}, ..., \mathbf{v_n}$, and $m > n$, then $\mathbf{u_1}, ..., \mathbf{u_m}$ are linearly dependent.* [1]

1.3 Basis

Definition 5. For a set S in a vector space $(V, +, \cdot)$, the vectors $\mathbf{v_1}, ..., \mathbf{v_n} \in S$ are a (finite) basis of S, iff

1. $\mathbf{v_1}, ..., \mathbf{v_n}$ are linearly independent, and
2. All vectors in S can be represented by linear combinations of $\mathbf{v_1}, ..., \mathbf{v_n}$.

Remark 1. because of condition (1), we know that the representation in condition (2) is unique.

Remark 2. the basis of S is not unique. However, two bases must have the same number of vectors. To see this, suppose $\mathbf{u_1}, ..., \mathbf{u_m}$ and $\mathbf{v_1}, ..., \mathbf{v_n}$ are both bases of S, and $m < n$. Then $\mathbf{u_1}, ..., \mathbf{u_m}$ can represent $\mathbf{v_1}, ..., \mathbf{v_n}$, and therefore $\mathbf{v_1}, ..., \mathbf{v_n}$ are linearly dependent by Theorem 1. This contradicts the assumption that $\mathbf{v_1}, ..., \mathbf{v_n}$ are a basis of S.

So, it is without ambiguity to define the rank of S   V as the number of vectors in its basis, denoted as $Rank(S)$.

Proposition 3. *In vector space $(V, +, \cdot)$, $\mathbf{v_1}, ..., \mathbf{v_n}$ are a basis of $S \subset V$ iff they have the maximum number of linearly independent vectors in S.*

In vector space (V,+, ·), $v_1, ..., v_n$ are a basis of $S \subset V$ iff they have the minimum number of vectors in S that can represent all vectors in S as their linear combinations.

Proposition 4. *In vector space $(V, +, \cdot)$, suppose $\mathbf{v_1}, ..., \mathbf{v_n}$ are a basis. then $Span(S) = Span(\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_n})$ of $S \subset V$.*

[1] *Shortly put, if "more" can be represented by "less", then "more" must be linearly dependent. d*

If we let $S = V$, we can talk about the basis and the rank of the whole vector space V . Then rank of the whole space V is usually called the dimension of the vector space V and denoted as dim V . If dim$V = n$, then we say that the vector space is n-dimensional. Clearly, any set of more than n vectors in an n-dimensional vector space must be linearly dependent. For example, the space $\mathbb{P}_n$ of all polynomials of degree at most n, consisting of all polynomials of the form $p(t) = a_0 + a_1 t + ... + a_n t^n$; is an $n+1$-dimensional vector space, for which one basis is $1, t, ..., t^n$.
It is also possible that a vector space V does not have a finite basis, in which case we say that V is infinite-dimensional.
In an n-dimensional vector space, it can be shown that $v_1, ..., v_n$ are a basis of V if they are linearly independent, or they can represent all vectors in V . The canonical basis of the vector space $\mathbb{R}^n$ is $e_1, ..., e_n$, where $e_1 := (1, 0, 0, ..., 0)$, $e_2 := (0, 1, 0, ..., 0)$, and so forth. Therefore dim$\mathbb{R}^n = n$.

1.4 Inner Products, Norms, and Metrics

Let's define an inner product operator on a real vector space V, to give it more structures.
Let $(V, +, \cdot)$ be a vector space. The 4-tuple $(V, +, \cdot, \langle \cdot, \cdot \rangle)$ is an inner product space iff the inner product operator $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ satisfies the following properties:

1. Commutativity: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$ for any $\mathbf{u}, \mathbf{v} \in V$,
2. Linearity: $\langle \lambda_1 \mathbf{u_1} + \lambda_2 \mathbf{u_2}, \mathbf{v} \rangle = \lambda_1 \langle \mathbf{u}_1, \mathbf{v} \rangle + \lambda_2 \langle \mathbf{u}_2, \mathbf{v} \rangle$ for any $\lambda_1, \lambda_2 \in \mathbb{R}$ and $\mathbf{u_1}, \mathbf{u_2}, \mathbf{v} \in V$, and
3. Positive definiteness: $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ for any $\mathbf{u} \in V$, and equality holds iff $\mathbf{u} = 0$.

Note that linearity also implies $\langle \mathbf{v}, \mathbf{0} \rangle = 0$ for any $\mathbf{v} \in \mathbf{V}$, because

$$\langle v, 0 \rangle = \langle v, 0v \rangle = 0 \cdot \langle \mathbf{v}, \mathbf{u} \rangle = 0$$

where $\mathbf{u}$ is an arbitrary vector in V.
In an inner product space $(V, +, \cdot, \langle \cdot, \cdot \rangle)$, two vectors $\mathbf{v}$ and $\mathbf{u}$ are said to be orthogonal iff $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.
A leading example of inner products is the dot product defined on $\mathbb{R}^n$. The dot product of two vectors $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ in $\mathbb{R}^n$ is defined as

$$\mathbf{x} \cdot \mathbf{y} := \sum_{i-1}^{n} x_i y_i$$

Notice that the dot product above defined for two vectors is different from the scalar multiplication, which is defined for a scalar and a vector, although we usually use the same notation $\cdot$. It is straightforward to verify that the dot product satisfies our requirements on inner products.
An inner product induces a **norm** $|| \cdot || : \cdot V \to \mathbb{R}_+$ by

$$||v|| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

Using the properties of inner product, it is straightforward to show that (1) $||v|| = 0$ iff $\mathbf{v} = \mathbf{0}$, and (2) $||\lambda v|| = |\lambda| \cdot ||v||$.
In $\mathbb{R}^n$, the norm induced by the dot product

$$\|x\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2}$$

is called the Euclidean norm, or L_2 norm.

Theorem 2. *(Cauchy-Schwarz Inequality). In an inner product space $(V, +, \cdot, \langle \cdot, \cdot \rangle)$, we have*

$$\|\mathbf{u}\| \, \|\mathbf{v}\| \geq |\langle \mathbf{u}, \mathbf{v} \rangle|$$

for any $\mathbf{u}, \mathbf{v} \in V$.

Proof. If $\mathbf{u} = 0$, the inequality holds trivially.
Now consider the case where $\mathbf{u} \neq 0$. First, I claim that the vectors $\lambda \mathbf{u}$ and $\mathbf{v} - \lambda \mathbf{u}$ are orthogonal, where the real number λ is given by

$$\lambda := \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}||^2}$$

This is because

$$\begin{aligned} \langle \lambda \mathbf{u}, \mathbf{v} - \lambda \mathbf{u} \rangle &= \lambda \langle \mathbf{u}, \mathbf{v} - \lambda \mathbf{u} \rangle = \lambda \left[\langle \mathbf{u}, \mathbf{v} \rangle - \lambda \langle \mathbf{u}, \mathbf{u} \rangle \right] \\ &= \lambda \left[\langle \mathbf{u}, \mathbf{v} \rangle - \lambda ||\mathbf{u}||^2 \right] = 0 \end{aligned}$$

Therefore, we have

$$\begin{aligned} ||v||^2 &= ||\lambda \mathbf{u} + (\mathbf{v} - \lambda \mathbf{u})||^2 \\ &= \langle \lambda \mathbf{u} + (\mathbf{v} - \lambda \mathbf{u}), \lambda \mathbf{u} + (\mathbf{v} - \lambda \mathbf{u}) \rangle \\ &= \langle \lambda \mathbf{u}, \lambda \mathbf{u} \rangle + 2 \langle \lambda \mathbf{u}, (\mathbf{v} - \lambda \mathbf{u} \rangle + \langle \mathbf{v} - \lambda \mathbf{u}, \mathbf{v} - \lambda \mathbf{u} \rangle \\ &= \lambda^2 ||u||^2 + ||\mathbf{v} - \lambda \mathbf{u}||^2 \end{aligned}$$

As a result, we have $||\mathbf{v}||^2 \geq \lambda^2 ||u||^2$, i.e.,

$$||\mathbf{v}||^2 \geq \left(\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}||^2} \right)^2 ||\mathbf{u}||^2$$

i.e., $||v||^2 ||\mathbf{u}||^2 \geq \langle \mathbf{u}, \mathbf{v} \rangle^2$, and therefore we have $||\mathbf{u}|| \; ||\mathbf{v}|| \geq | \langle \mathbf{u}, \mathbf{v} \rangle |$. □

Notice Cauchy-Schwarz inequality also tells us that any norm induced by an inner product satisfies the **triangle inequality**: $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + \|\mathbf{v}||$ for any $\mathbf{u}, \mathbf{v} \in V$.

Definition 6. Let $(V, +, \cdot)$ be a vector space. The 4-tuple $(V, +, \cdot, || \cdot ||)$ is a normed vector space iff the norm $|| \cdot || : V \to \mathbb{R}_+$ satisfies the following properties:
(1)$||\mathbf{v}|| = 0$ iff $\mathbf{v} = \mathbf{0}$, for any $v \in V$,
(2) $||\lambda \mathbf{v}|| = |\lambda| \cdot ||\mathbf{v}||$, for any $\lambda \in \mathbb{R}$ and $\mathbf{v} \in V$, and
(3) Triangle inequality: $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$, for any $\mathbf{u}, \mathbf{v} \in V$.

If we consider $\mathbb{R}^n$ endowed with the dot product as an inner space, then the dot product induces the L_2 norm, which in turn induces the Euclidean distance

$$d_2(\mathbf{x}, \mathbf{y}) \coloneqq \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Recall that we have shown that a valid inner product induces a valid norm, which in turn induces a valid metric. Because the dot product on $\mathbb{R}^n$ is a valid inner product, the Euclidean distance function d_2 is a valid metric; especially, it satisfies the triangle inequality required for metrics.

2 Matrices

An $m \times n$ matrix is an array with $m \geq 1$ rows and $n \geq 1$ **columns**:

$$A = (a_{ij})_{ij} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

where the number a_{ij} in the i^{th} row and j^{th} column is called the ij^{th}-entry or the ij^{th}-component. The set of matrices of size $m \times n$ is noted $\mathcal{M}_{mn}$.
We note A_i the i^{th} row of A and A^j the j^{th} column of A, so that:

$$A = (A^1 ... A^n) = \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix}$$

Here are some particular matrices:
• Real numbers can be seen as a 1×1 matrix.
• A vector of Rk can be seen as a $k \times 1$ matrix (a column vector) or a matrix of size $1 \times k$ (a row vector). By default a vector is seen as a column vector.
• Matrices with as many rows as columns $m = n$ are called square matrices.
• The zero matrix of $\mathcal{M}_{mn}$ is the matrix with all entries equal to zero.
• A square matrix A is diagonal if all its non-diagonal elements are zero: $a_{ij} = 0$ for all i, j such that $i \neq j$. We note $A = diag(a_{11}, ..., a_{nn})$.
• The unit matrix of size n is the square matrix of size n having all its components equal to zero except the diagonal components, equal to 1. It is noted I_n.
• A square matrix A is upper-triangular if all its elements below its diagonal are nil: $a_{ij} = 0$ for all $i > j$.
• A square matrix A is lower-triangular if all its elements above its diagonal are nil: $a_{ij} = 0$ for all $i < j$.

2.1 Operations on matrices

First, addition and scalar multiplications—the set of matrices Mmn is going to be a vector space. These two are defined on the setMmn of matrices of the same size m×n.
Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices, and $\lambda \in \mathbb{R}$.
(1) The sum $A + B$ is the matrix whose ij-entry is $a_{ij} + b_{ij}$.
(2) The scalar multiplication of A by λ, λA is the matrix whose ij^{th}-entry is λa_{ij}.
The matrix multiplication is defined over matrices of different sizes, although sizes need to be conformable: the product AB is only defined for matrices such that the number of columns of A is equal to the number of rows of B.

Definition 7. Let $A = (a_{ij})$ be an $m \times n$ matrix and $B = (b_{ij})$ be an $n \times s$ matrix. Their product AB is the $n \times s$ matrix whose ij^{th}-entry is:

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

In fact, matrix multiplication should be interpreted as a linear mapping. Recall that in an $m \times n$ matrix with components in $\mathbb{R}$, each column can be viewed as a vector in $\mathbb{R}^m$, and each row can be viewed as a vector in $\mathbb{R}^n$. If we left-multiply matrix A by another matrix C, each row of the product CA is a linear combination of the rows of A; therefore, left-multiplying a matrix can be interpreted as a row transformation. If we right-multiply A by another matrix C, each column of the product AC is a linear combination of the columns of A; therefore, right-multiplying a matrix should be interpreted as a column transformation. The following properties of matrix multiplication are easy to verify from the definition.
Provided conformable matrices:

- The unit matrix is the neutral element of matrix multiplication: if A is $m \times n$, then $I_m A = A I_n = A$.
- The zero matrix is absorbant: $A0 = 0A = 0$.
- The multiplications is distributive wrt. the addition: $A(B + C) = AB + AC$ and $(B + C)A = BA + CA$.
- The multiplication is associative: $A(BC) = (AB)C$.
- $A(\lambda B) = \lambda(AB)$. But be careful that, contrary to the multiplication on real numbers, the matrix multiplication is in general NOT commutative: in general $AB \neq BA$. Here is a counterexample:

  $$A = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}, \ B = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$$

  $AB = 0$ does NOT imply that either A or B is zero, as it does for the multiplication of reals.

A square matrix of size n is invertible (or non-singular) iff there exists a matrix B such that $AB = BA = I_n$. Provided existence, the inverse is unique and noted A^{-1}.

To prove the uniqueness of the inverse, assume that B and C are two inverses of A. Then $B = BI_n = B(AC) = I_nC = C$, which proves that all inverses of A are equal. Obviously, if B is the inverse of A, then A is the inverse of A, so the inverse of the inverse of A is A itself: $(A^{-1})^{-1} = A$. Besides:

Proposition 5. *If $A, B \in \mathcal{M}_{nn}$ are invertible, then so is their product AB and:*

$$(AB)^{-1} = B^{-1}A^{-1}$$

For square matrices, it is also possible to define the repeated products, or powers of a square matrix A.

Definition 8. $A^k = A, ..., A$ taken k times. By definition, $A^0 = I_n$.

We say that a matrix A is idempotent if $A^2 = A$. We say that a matrix A is nilpotent if $A^k = 0$ for some integer k.
Another operation on matrices, although less essential, is the transpose; it takes simply one argument—a matrix—and returns another matrix.

Definition 9. Let $A = (a_{ij})$ be a matrix. The transpose A' (or A^T) of A is the matrix obtained by changing its rows into its columns (and vice versa): $A^T = (a_{ji})$.

- $(\lambda A)^T = \lambda A^T$.
- Transpose of the sum: $(A + B)^T = A^T + B^T$.
- Transpose of the product: $(AB)^T = B^T A^T$.
- $(A^{-1})^T = (A^T)^{-1}$ (provided the inverse exists).

Definition 10. A square matrix A is symmetric iff it is equal to its transpose $A' = A$.

If A is an $m \times n$ real matrix, we can see its n columns $A^1, ..., A^n$ as n vectors of $\mathbb{R}^m$. Conversely, if $A^1, ..., A^n$ are n vectors of $\mathbb{R}^m$, we can see them as the $m \times n$ matrix whose columns are the A^j. For instance, note this very useful way to write a linear combination of the vectors A^j using matrix multiplication (just check the equality entry by entry):

$$\lambda_1 A^1 + ... + \lambda_n A^n = A\lambda, \ \textit{where } \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

Since we can look at matrices as a family of vectors, we can also consider the vector space that these vectors span:

Definition 11. If $A = (A^1, ..., A^n)$ is an $m \times n$ matrix, we call the space $Span(A^1, ..., A^n)$ spanned by the columns of A the column space (or image, noted $Im(A)$) of the matrix A.

We can do with rows what we have done with columns. We can see the n rows of an $m \times n$ matrix as m vectors of Rn. We call the space $Span(A_1, ..., A_m)$ spanned by the row vectors of A the row space of matrix A and its rank the row rank of the matrix. However, the row space is not as much useful as the column space, and:

Proposition 6. *The row rank of a matrix is equal to its column rank.*

Because the rows of the product matrix AB are linear combinations of rows of B, the basis of the rows of B can represent rows of AB, and so we have $Rank(AB) \leq Rank(B)$. Because the columns of AB are linear combinations of columns of A, the basis of columns of A can represent columns of AB, and so we have $Rank(AB) \leq Rank(A)$. As a result, we always have

$$Rank(AB) \leq \min\{Rank(A), Rank(B)\}$$

A consequence is that the rank of an m $\times$ n matrix is always smaller than both n and m.

Proposition 7. *A square matrix A of size n is invertible iff $rank(A) = n$.*

Definition 12. Let $A = (a_{ij})$ be a square matrix of size n. The trace of A, noted $tr(A)$, is $tr(A) = \sum_{i=1}^{n} a_{ii}$.

- The trace is linear: $tr(\ \lambda A) = \ \lambda \cdot tr(A)$ and $tr(A+B) = tr(A) + tr(B)$.
- A matrix and its transpose have the same trace: $tr(A') = tr(A)$.
- If AB and BA are square (but not necessarily A and B), $tr(AB) = tr(BA)$.

Definition 13. For a square matrix A, its **determinant**, denoted as $det(A)$, is an element defined inductively in the following way:
(1) For a 1×1 matrix$A = a_{11}$, define its determinant as $det(A) := a_{11}$.
(2) For an $n \times n$ matrix where $n \geq 2$, define its determinant as

$$det(A) := \sum_{j=1}^{n} (-1)^{1+j} a_{1j} det(A_{-1,-j})$$

where $A_{-i,-j}$ is the matrix A with the i-th row and j-th column eliminated

According to the definition above, for a 2 $\times$ 2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Its determinant is det $A = a_{11}a_{22} - a_{12}a_{21}$.
For a 3×3 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & s_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Its determinant is

$$\begin{aligned} det(A) &= a_{11} det\left(\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}\right) - a_{12} det\left(\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}\right) + a_{13} det\left(\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\right) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \end{aligned}$$

In the inductive definition of determinants above, the induction formula

$$det(A) := \sum_{j=1}^{n} (-1)^{1+j} a_{1j} det(A_{-1,-j})$$

is also called the cofactor expansion of A along the first row. The (i, j)-th cofactor of a square matrix A, denoted as A_{ij}, is defined as

$$A_{ij} := (-1)^{i+j} det(A_{-i,-j})$$

and so the cofactor expansion of A along the first row can be rewritten as

$$det(A) := \sum_{j=1}^{n} a_{1j} A_{1j}$$

In fact, we can equivalently define determinants by expanding along any row or column, i.e.

$$det(A) := \sum_{j=1}^{n} a_{ij} A_{ij}$$

for any arbitrary row i, or

$$det(A) := \sum_{i=1}^{n} a_{ij} A_{ij}$$

for any arbitrary column j. Let's admit the equivalence of different ways of expansion without providing a proof. Using this equivalence of expanding along a row or a column, we can show that $det(A^T) = det(A)$.

Theorem 3. *Let n be an integer and consider the determinant function, taking n vectors of $\mathbb{R}^n$ as argument det: $\mathbb{R}^n \times ... \times \mathbb{R}^n \to \mathbb{R}$.*

1. $det(I_n) = 1$.
2. The determinant of a triangular (this includes diagonal) matrix $A = (a_{ij})$ is the product of its diagonal elements $det(A) = \prod_{i=1}^{n} a_{ii}$.
3. Multilinearity: det is linear with respect to each of its argument:

   $$\forall k = 1...n, \det(A^1, ..., \lambda A^k + \mu A^{k'}, ..., A^n) = \lambda det(A^1, ..., A^k, ..., A^n) + \mu det(A^1, ..., A^{k'}, ..., A^n)$$

4. If any two columns of A are equal, then $det(A) = 0$.
5. Antisymmetry: If two columns of A are interchanged, then the determinant changes by a sign.
6. If one adds a scalar multiple of one columns to another then the determinant does not change.

Theorem 4. *..*

1. *(1) A matrix and its transpose have the same determinant: $det(A') = det(A)$. (Equivalently, n vectors $A^1, ..., A^n$ of $\mathbb{R}^n$ are linearly independent iff $det(A^1, ..., A^n) \neq 0$.)*
2. *(2) A square matrix A is invertible iff $det(A) \neq 0$.*
3. *(3) For two $n \times n$ matrices, we have $det(AB) = det(A)det(B)$.*

The Theorem above implies that $det(A^{-1}) = (det(A))^{-1}$ for an invertible matrix A.

3 Systems of Linear Equations

Consider the following system of linear equations in x:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m \end{cases}$$

where a_{ij}'s and b_i's are all elements of $\mathbb{R}$, and the unknowns $x_1, ..., x_n$ also take values in $\mathbb{R}$. We can rewrite the system of linear equations in a compact way by

$$A\mathbf{x} = \mathbf{b}$$

where $A = (a_{ij})$ is the $m \times n$ matrix, $\mathbf{b} = (b_1, ..., b_m)^T$, and $\mathbf{x} = (x_1, ..., x_n)^T$.
If we view $\mathbf{x}$ as a column transformation of the columns of A, then the equation asks us to find ways to represent the vector $\mathbf{b} \in \mathbb{R}^m$ as a linear combination of the columns of A. To clearly see this, write matrix A as $A = (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n)$, where $\mathbf{a}_i$ is the ith column of A, we have $X = (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n)(x_1, ..., x_n)^T = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n$.

Proposition 8. *The system of equations $A\mathbf{x} = \mathbf{b}$ (A is an $m \times n$ matrix over $\mathbb{R}$) has a solution iff*
$Rank([A|b]) = Rank(A)$. When the system has a solution,

1. *the solution is unique iff the columns of A are linearly independent, i.e. $Rank(A) = n$.*
2. *the system has infinitely many solutions iff $Rank(A) < n$.*

Proposition 9. *(General solution of a linear equation). Let a vector x^* satisfy the equation $Ax^* = \mathbf{b}$, and let $H := \{\mathbf{z} : \mathbf{Az} = \mathbf{0}\}$. Then the set $j\{x = x^* + x_h : x_h \in H\}$ is the set of all solutions of the equation $\mathbf{Ax} = \mathbf{b}$.*

The system of linear equations can be solved by hand using Gauss-Jordan elimination, which is essentially row operations of the matrix $[A|b]$. You may refer to a standard linear algebra textbook for details.
As a special case, when $m = n$ and the square matrix A is invertible, the unique solution is clearly $\mathbf{x}^* = A^{-1}b$. We also have an explicit formula for $x^* = A^{-1}b$, which is known as Cramer's rule. Let's state it without proof.

Theorem 5. *(Cramer's Rule). Let A be an $n \times n$ invertible matrix and b be an $n \times 1$ column vector. The i-th entry of the $n \times 1$ column vector $\mathbf{x}^* := \mathbf{A^{-1}b}$ can be calculated as*

$$x_i^* = \frac{det(A_i)}{det(A)}$$

for each i, where A_i is the $n \times n$ matrix formed by replacing the i-th column of A by **b** *and leaving the other columns unchanged.*

However, calculating determinants is numerically difficult when the size of the matrices is large, since the determinant of an $n \times n$ matrix has $n!$ terms. So Cramer's rule may not be as useful as it seems.

4 Eigenvalues, Eigenvectors, and Diagonalization

Let A be an $n \times n$ matrix over $\mathbb{C}$. A scalar $\lambda \in \mathbb{C}$ is said to be an eigenvalue of A iff $\exists x \in \mathbb{C}^n \backslash \{0\}$ s.t. $Ax = \lambda x$. A vector $x \in \mathbb{C}^n \backslash \{0\}$ is said to be an eigenvector of A iff $\exists \lambda \in \mathbb{C}$ s.t. $\mathbf{Ax} = \lambda \mathbf{x}$.

Proposition 10. *$\lambda \in \mathbb{C}$ is an eigenvalue of A iff $det(\lambda I_n - A) = 0$.*

By definition, $\lambda \in \mathbb{C}$ is an eigenvalue of A iff $\mathbf{Ax} = \lambda \mathbf{x}$ has a nonzero solution. This is equivalent to $(A - \lambda I_n)\mathbf{x} = 0$ having a nonzero solution, which is in turn equivalent to the columns of the matrix $\lambda I_n - A$ being linearly dependent, which is in turn equivalent to $(\lambda I_n - A) = 0$.
In the determinant of the matrix

$$\lambda I_n - A = \begin{bmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{bmatrix}$$

the diagonal contributes a term λ^n, and all other terms have a degree no more than $n-2$. Therefore, $det(\lambda I_n - A)$ is a polynomial of λ of degree n. The polynomial $P_A(\lambda) := det(\lambda I_n - A)$ is also called the characteristic polynomial of A. By construction, $\lambda \in \mathbb{C}$ is an eigenvalue of A iff $P_A(\lambda) = 0$.

Theorem 6. *(Fundamental Theorem of Algebra). Let $P : \mathbb{C} \to \mathbb{C}$ be a polynomial of degree n, i.e. $P(\lambda) = c_n\lambda^n + c_{n-1}\lambda^n + ... + c_1\lambda + c_0$, where $c_k \in \mathbb{C}$ for any $k = 0, 1, ..., n$ and $c_n \neq 0$. Then P has exactly n roots in C, counted with multiplicity. That is, there exists $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{C}$ s.t.*

$$P(\lambda) = c_n(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$$

Therefore, we can obtain all eigenvalues of A by setting the characteristic polynomial of A to 0 and solving for all its roots.

Proposition 11. *Let A be an $n \times n$ matrix over $\mathbb{C}$ and $\lambda_1, \lambda_2 ..., \lambda_n \in \mathbb{C}$ are eigenvalues of A. The the characteristic function polynomial of A:*

$$P_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$$

Let A be an $n \times n$ matrix over $\mathbb{C}$ and $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{C}$ are eigenvalues of A. Then $det(A) = \lambda_1 \lambda_2 ... \lambda_n$

Proof. $\lambda = 0$ in $P_A(\lambda) = det(\lambda I_n - A)$, we have $det(-A) = (-1)^n = (0-\lambda_1)(0-\lambda_2)...(0-\lambda_n) = (-1)^n \lambda_1 \lambda_2 ... \lambda_n$. $\square$

Suppose we have established that $\lambda \in \mathbb{C}$ is an eigenvalue of A. Then we can obtain the set of all eigenvectors associated with λ by solving for all nonzero solutions in $\mathbb{C}^n$ to the system of linear equations $(A - \lambda I_n)x = 0$.

Definition 14. Let A be an $n \times n$ matrix over $\mathbb{C}$. The matrix A is diagonalizable in $\mathbb{C}$ iff there exists an $n \times n$ invertible matrix P over $\mathbb{C}$ and an $n \times n$ diagonal matrix Λ over $\mathbb{C}$ s.t. $P^{-1}AP = \Lambda$. The matrix A is **diagonalizable** in $\mathbb{R}$ iff there exists an $n \times n$ invertible real matrix P and an $n \times n$ diagonal real matrix Λ s.t. $P^{-1}AP = \Lambda$.

Intuitively, the matrix A is diagonalizable iff we can find invertible matrix P s.t. we can transform A into some diagonal matrix by left-multiplying A by P^{-1} and right-multiplying A by P.
When a matrix A is diagonalizable, i.e. $P^{-1}AP = \Lambda$, we have

$$A = (PP^{-1})A(PP^{-1}) = P(P^{-1}AP)P^{-1} = P\Lambda P^{-1}$$

and so the matrix A can be decomposed as $P\Lambda P^{-1}$.
Not all matrices are diagonalizable. For example, consider the matrix

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

It is straightforward to see that the two eigenvalues of A are both 0. So if A can be diagonalized as Λ under P, i.e. $P^{-1}AP = \Lambda$, the diagonal matrix Λ must be the 2×2 zero matrix. Then we have $A = P\Lambda P^{-1} = 0$. Contradiction. The next proposition establishes a necessary and sufficient characterization of diagonalizable matrices.

Proposition 12. *Let A be an $n \times n$ matrix over $\mathbb{C}$. Then A is diagonalizable in $\mathbb{C}$ iff A has n linearly independent eigenvectors.*

Proposition 13. *Let A be an $n \times n$ matrix over $\mathbb{C}$ with n distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{C}$. Let $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n} \in \mathbb{C}^n \backslash \{0\}$ be the corresponding eigenvectors. Then $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}$ are linearly independent.*

Theorem 7. *Let A be an $n \times n$ matrix over $\mathbb{C}$ with n distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{C}$. Then A is diagonalizable in $\mathbb{C}$.*

If the eigenvalues of A are not distinct, we don't know whether A is diagonalizable or not.
A matrix over field R is called a real matrix. An $n \times n$ matrix A is symmetric iff $a_{ij} = a_{ji}$ for any i and j. An $n \times n$ matrix A over F is orthogonal iff $A^T A = I_n$. Clearly, the condition $A^T A = I_n$ means that the columns of A are pairwise orthogonal (w.r.t. dot product) and each have a norm of 1. The condition$A^T A = I_n$ also implies that A is invertible and $A^{-1} = A^T$, and the columns of A are a basis of F^n. The next theorem states that a real symmetric matrix is always diagonalizable in R, i.e. there exists a real invertible matrix P and a real diagonal matrix s.t. $P^{-1}AP = \Lambda$.

Theorem 8. *Let A be an $n \times n$ real symmetric matrix. Then all its eigenvalues are real, and there exists a real orthogonal matrix P and a real diagonal matrix s.t. $P^{-1}AP = P^TAP = \Lambda$.*

For an economist, the motivation for studying eigenvalues, eigenvectors, and diagonalization is their applications in dynamic models. Consider a linear dynamic system $x_t = Ax_{t-1}$, where x_t is an n-dimensional real vector and A is an $n \times n$ real matrix. Clearly we have $x_t = A^t x_0$. When Λ^t is large, it is difficult to analyze the behavior of x_t since calculating A^t is difficult. With the help of diagonalization, however, $A^t = (PAP^{-1})^t = P\Lambda^t P^{-1}$, where Λ^t is easy to calculate since Λ is diagonal. In fact, if all eigenvalues of A have a modulus strictly less than 1, then $\Lambda^t \to 0$ (the $n \times n$ zero matrix), and so $\mathbf{x}_t = P\Lambda^t P^{-1}\mathbf{x}_0 \to \mathbf{0}$ (the n-dimensional zero vector).
If the dynamic system is not linear, it is a standard practice in macro to log-linearize the dynamic system around its steady state, which is essentially approximating a nonlinear system using a linear system. Then our discussion on linear dynamic systems above applies.

5 Quadratic Forms

A quadratic form on $\mathbb{R}^n$ is a function $Q : \mathbb{R}^n \to \mathbb{R}$ that can be represented by

$$\begin{aligned}
Q(x) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \\
&= a_{11}x_1^2 + a_{22}x_2^2 + ... + a_{nn}x_n^2 \\
&+ (a_{12} + a_{21})x_1x_2 + (a_{13} + a_{31})x_1x_3 \\
&+ ... + (a_{n-1,n} + a_{n,n-1})x_{n-1}x_n
\end{aligned}$$

where a_{ij}'s are real coefficients.
Notice that we can write a quadratic form in a compact way using matrix multiplication. The quadratic form Q(x) defined above is equal to x^TAx, where $A = (a_{ij})$ is the $n \times n$ matrix whose elements are the coefficients of the quadratic form, and x is considered as a column vector.
The way to represent a quadratic form Q using a matrix A is not unique, since if the matrix A represents the quadratic form Q, then the matrix $A + B$ also represents Q for any anti-symmetric matrix B (i.e. $b_{ij} = -b_{ji}$ for any i, j). However, each quadratic form Q can be represented by a unique symmetric matrix A.

Definition 15. Let A be an $n \times n$ real symmetric matrix. The matrix A, or the quadratic form $Q(x) := x^TAx$ that is represented by A, is said to be
(1) positive definite, iff $x^TAx > 0$ for any $x \in \mathbb{R}^n \backslash \{0\}$;
(2) negative definite, iff $x^TAx < 0$ for any $x \in \mathbb{R}^n \backslash \{0\}$;
(3) positive semi-definite, iff $x^TAx \geq 0$ for any $x \in \mathbb{R}^n$;
(4) negative semi-definite, iff $x^TAx \leq 0$ for any $x \in \mathbb{R}^n$;
(5) indefinite, iff $\exists \mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$ s.t.$x^TAx > 0$ and $x'^TAx' < 0$. The next theorem provides a necessary and sufficient characterization of positive/negative (semi-)definite matrices using eigenvalues.

Theorem 9. *Let A be an $n \times n$ real symmetric matrix. The matrix A is*

1. *positive definite iff all its eigenvalues are positive;*
2. *negative definite iff all its eigenvalues are negative;*
3. *positive semi-definite iff all its eigenvalues are non-negative;*
4. *negative semi-definite iff all its eigenvalues are non-positive;*
5. *indefinite iff it has both positive and negative eigenvalues.*

This theorem easily follows, which allows us to find orthogonal P and diagonal A s.t. $P^T AP = \Lambda$, since we have

$$\mathbf{x^T A x} = \mathbf{x^T (P \Lambda P^T) x} = \mathbf{(P^T x)^T \Lambda (P^T x)} = \mathbf{y^T \Lambda y} = \sum_{i=1}^{n} \lambda_i y_i^2$$

where $\mathbf{y} \coloneqq \mathbf{P^T x}$, and λ_i's are entries on the diagonal of Λ.

Theorem 10. *Let A be an $n \times n$ real symmetric matrix. Then A is positive definite iff there exists a real diagonal matrix D with positive entries on its diagonal and a real lower triangle matrix L with all 1's on its diagonal, s.t. $A = LDL^T$.*

Theorem 11. *Let A be an $n \times n$ real symmetric matrix. Then A is positive definite iff there exists a real lower triangle matrix P with all positive entries on its diagonal s.t. $A = PP^T$.*