Lecture 1: Fundamentals of Linear Algebra and Matrix Analysis

Tianpei Xie

Sep. 1st., 2022

Contents

1 Basics in Linear Algebra

- A ***vector space*** over a field F is a set V together with two operations, the (vector) addition $+ : V \times V \to V$ and scale multiplication $\cdot : \mathbb{R} \times V \to V$, that satisfy the eight axioms listed below: for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$, $\alpha, \beta \in F$,

  1. The *associativity* of addition: $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}$;
  2. The *commutativity* of addition: $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$;
  3. The *identity* of addition: $\exists\, \mathbf{0} \in V$ such that $\mathbf{0} + \boldsymbol{x} = \boldsymbol{x}$;
  4. The *inverse* of addition: $\forall\, \boldsymbol{x} \in V$, $\exists\, -\boldsymbol{x} \in V$, so that $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$;
  5. *Compatibility* of scalar multiplication with field multiplication: $\alpha(\beta \cdot \boldsymbol{x}) = (\alpha\beta) \cdot \boldsymbol{x}$;
  6. The *identity* of scalar multiplication: $\exists\, 1 \in F$, such that $1 \cdot \boldsymbol{x} = \boldsymbol{x}$;
  7. The *distributivity* of scalar multiplication with respect to vector addition: $\alpha \cdot (\boldsymbol{x} + \boldsymbol{y}) = \alpha \cdot \boldsymbol{x} + \alpha \cdot \boldsymbol{y}$;
  8. The *distributivity* of scalar multiplication with respect to field addition: $(\alpha + \beta) \cdot \boldsymbol{x} = \alpha \cdot \boldsymbol{x} + \beta \cdot \boldsymbol{x}$.

  Elements of V are commonly called *vectors*. Elements of F are commonly called *scalars*.

- A vector space X endowed with a topology is called a *topological vector space*, denoted as $(X, \mathscr{T})$, if the addition $+ : X \times X \to X$ and scale multiplication $\cdot : \mathbb{R} \times X \to X$ are continuous.

- A ***subspace*** $S \subset V$ over a field F is, by itself, a vector space over F that *is closed* under the same operations of the vector addition and scalar multiplication as in V.

  The subsets $\{0\}$ and V are always subspaces of V , so they are often called *trivial subspaces*; a subspace of V is said to be *nontrivial* if it is different from both $\{0\}$ and V. We call $\{0\}$ the *zero vector space*. A subspace of V is said to be a *proper subspace* if it is not equal to V.

- A ***linear combination*** of vectors in a vector space V over a field F is any expression of the form $a_1\boldsymbol{v}_1 + \ldots + a_k\boldsymbol{v}_k$ in which k is a *positive integer*, $a_1, \ldots, a_k \in F$, and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in V$.

  A linear combination is *trival* if $a_1 = \ldots = a_k = 0$; otherwise, it is *nontrivial*. A linear combination is by definition a sum of *finitely many elements* of a vector space.

- The *span* of a nonempty subset S of V, $\mathrm{span}(S)$, consists of *all linear combinations* of finitely many vectors in S.

- Let S_1 and S_2 be subspaces of a vector space over a field F. The *sum* of S_1 and S_2 is the *subspace*

  $$S_1 + S_2 = \mathrm{span}\{S_1 \cap S_2\} = \{\boldsymbol{x} + \boldsymbol{y} : \boldsymbol{x} \in S_1, \boldsymbol{y} \in S_2\}$$

  If $S_1 \cap S_2 = \{0\}$, we say that the sum of S_1 and S_2 is a *direct sum* and write it as $S_1 \oplus S_2$; every $\boldsymbol{z} \in S_1 \oplus S_2$ can be written as $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$ with $\boldsymbol{x} \in S_1$ and $\boldsymbol{y} \in S_2$ in one and *only one way*.

1.1 Linear indepenence

- We say that a finite list of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ in a vector space V over a field F is ***linearly dependent*** *if and only if* there are scalars $a_1, \ldots, a_k \in F$, *not all* zero, such that $a_1\boldsymbol{v}_1 + \ldots + a_k\boldsymbol{v}_k = \mathbf{0}$. Thus, a list of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ is *linearly dependent* if and only if some *nontrivial* linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ is the zero vector. A list of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ is said to have *length* k.

- A finite list of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ in a vector space V over a field F is ***linearly independent*** if and only if $a_1\boldsymbol{v}_1 + \ldots + a_k\boldsymbol{v}_k = \mathbf{0} \Leftrightarrow a_1 = \ldots = a_k = 0$.

- A list of vectors are linearly independent if and only if *every finite sublist* is linearly independent. Any list of vectors that *contains the zero* vector is linearly dependent.

- The ***cardinality*** of a finite set is the number of its (necessarily distinct) elements. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ are linearly independent, then the cardinality of the set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is k, i.e. there is no element that is identical to others.

- A set S of vectors is said to be linearly independent if *every finite list of distinct vectors* in S is linearly independent; S is said to be linearly dependent if some finite list of distinct vectors in S is linearly dependent.

- A *linearly independent* list of vectors in a vector space V whose span is V is a ***basis*** for V. Each element of V can be represented as a linear combination of vectors in a basis in **one *unique* way**.

$$V = \mathrm{span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}) = \{\boldsymbol{v} \in V : \boldsymbol{v} = a_1\boldsymbol{v}_1 + \ldots + a_k\boldsymbol{v}_k, a_i \in F\}$$

  Note that removing one basis vector will result is some $\boldsymbol{v}$ that cannot be linear represented by the rest.

- A linearly independent list of vectors $(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ in V is a basis of V if and only if no list of vectors that properly *contains it* is linearly independent, i.e. it is a maximum list of independent vectors.

- A list of vectors that *spans* V is a basis for V if and only if *no proper sublist* of it spans V. The empty list is a basis for the zero vector space.

- Any *linearly independent* list of vectors in a vector space V may be *extended*, perhaps in more than one way, to a *basis* of V.

1.2 Dimensions

- If there is a positive integer n such that a *basis* of the vector space V contains exactly n vectors, then *every basis* of V consists of exactly n vectors; this common cardinality of bases is the ***dimension*** of the vector space V and is denoted by $\dim(V)$. Here V is *finite-dimensional.*

  In the *infinite-dimensional* case, there is a one-to-one correspondence between the elements of any two bases.

- Let V be a finite-dimensional vector space and let S_1 and S_2 be two given subspaces of V.

The ***subspace intersection lemma*** is

$$\dim(S_1) + \dim(S_2) = \dim(S_1 \cap S_2) + \dim(S_1 + S_2) \tag{1}$$

The following inequality is true

$$\dim(S_1 \cap S_2) \geq \dim(S_1) + \dim(S_2) - \dim(V) \tag{2}$$

reveals the useful fact that if $\delta = \dim(S_1) + \dim(S_2) - \dim(V) \geq 1$, then the subspace $S_1 \cap S_2$ has dimension at least δ, and hence it contains δ linearly independent vectors, namely, any δ elements of a *basis* of $S_1 \cap S_2$.

This statement can be extended. If $S_1, \ldots, S_k$ are subspaces of V, and if $\delta = \dim(S_1) + \ldots + \dim(S_k) - (k-1)\dim(V) \geq 1$, then

$$\dim(S_1 \cap \ldots \cap S_k) \geq \delta \tag{3}$$

1.3 Isomorphism

- If U and V are vector spaces over the same scalar field F, and if $f : U \to V$ is an *invertible* function such that $f(a\boldsymbol{x} + b\boldsymbol{y}) = af(\boldsymbol{x}) + bf(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in U$ and all $a, b \in F$, then f is said to be an ***isomorphism*** and U and V are said to be ***isomorphic*** ("*same structure*"). Isomorphism is a *bijective* (one-to-one and onto) mapping that *preserve* linear operations.
- Two *finite-dimensional* vector spaces over the same field are *isomorphic* if and only if they have *the same dimension*.
- Any n-dimensional vector space over F is isomorphic to F^n.
- Specifically, if V is an n-dimensional vector space over a field F with specified basis $B = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, then, since any element $\boldsymbol{x} \in V$ may be written uniquely as $\boldsymbol{x} = a_1\boldsymbol{x}_1 + \ldots + a_n\boldsymbol{x}_n$ in which each $a_i \in F$, we may identify $\boldsymbol{x}$ with the n-vector $[\boldsymbol{x}]_B = [a_1, \ldots, a_n]^T$. For any basis B, the mapping $\boldsymbol{x} \to [\boldsymbol{x}]_B$ is an *isomorphism* between V and F^n. $[\boldsymbol{x}]_B$ is referred as the ***cooridnate*** of $\boldsymbol{x}$ in V. B forms a *coordinate system* of V.

2 Basics in Matrix

- A ***matrix*** is an m-by-n array of scalars from a field F. If $m = n$, the matrix is said to be *square*. The set of *all* m-by-n matrices over F is denoted by $M_{m,n}(F)$, and $M_{n,n}(F)$ is often denoted by $M_n(F)$.

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & \ldots & a_{1,n} \\ \ldots & \ldots & \ldots \\ a_{m,1} & \ldots & a_{m,n} \end{bmatrix} := [a_{i,j}]_{m \times n}$$

- The vector spaces $M_{n,1}(F)$ and F^n are identical.
- A *submatrix* of a given matrix is a *rectangular* array lying in specified *subsets* of the rows and columns of a given matrix.

- Suppose that $A = [ai, j] \in M_{n,m}(F)$. The ***main diagonal*** of A is the list of entries $a_{1,1}, a_{2,2}, \ldots, a_{q,q}$, in which $q = \min\{n, m\}$, denoted as $\operatorname{diag}(A) = [a_{i.i}]_{i=1}^{q} \in F^q$.

  The p-th *superdiagonal* of A is the list $a_{1,p+1}, a_{2,p+2}, \ldots, a_{k,p+k}$, in which $k = \min\{n, m-p\}$, $p = 0, 1, 2, \ldots, m-1$; the p-th *subdiagonal* of A is the list $ap+1, 1, ap+2, 2, \ldots, ap+l, l$, in which $l = \min\{n-p, m\}$, $p = 0, 1, 2, ..., n-1$.

2.1 Linear transformation

2.2 Matrix operations

2.3 Rank

2.4 Nonsingularity

3 The Euclidean inner product and norm

4 Partition set and matrices

4.1 Submatrices

- Let $\boldsymbol{A} \in M_{m,n}(F)$. For index sets $\alpha \subset \{1, \ldots, m\}$ and $\beta \subset \{1, \ldots, n\}$, we denote by $\boldsymbol{A}[\alpha, \beta]$ the (sub)matrix of entries that lie in the rows of $\boldsymbol{A}$ indexed by α and the columns indexed by β.

- If $\alpha = \beta$, the submatrix $\boldsymbol{A}[\alpha] = \boldsymbol{A}[\alpha, \alpha]$ is a ***principal submatrix*** of $\boldsymbol{A}$. An n-by-n matrix has $\binom{n}{k}$ distinct principal submatrices of size k

- For $\boldsymbol{A} \in M_n(F)$ and $k \subset \{1, \ldots, n\}$, $\boldsymbol{A}[\{1, \ldots, k\}]$ is a ***leading principal submatrix*** and $\boldsymbol{A}[\{k, \ldots, n\}]$ is a *trailing principal submatrix.*

- The ***determinant*** of an r-by-r submatrix of A is called a ***minor***; if we wish to indicate the size of the submatrix, we call its determinant *a minor of size* r.

  If the r-by-r submatrix is a principal submatrix, then its determinant is a ***principal minor*** (of size r); if the submatrix is a leading principal matrix, then its determinant is a ***leading principal minor***; if the submatrix is a trailing principal submatrix, then its determinant is a *trailing principal minor.*

- A ***signed*** minor, such as those appearing in the Laplace expansion $[(-1)^{i+j} \det \boldsymbol{A}_{i,j}]$ is called a ***cofactor***; if we wish to indicate the size of the submatrix, we call its signed determinant a *cofactor of size* r.

- Suppose that $\boldsymbol{A} \in M_n(F)$ and $\operatorname{rank}(\boldsymbol{A}) = r$. We say that $\boldsymbol{A}$ is ***rank principal*** if it has a nonsingular r-by-r principal submatrix.

  If there is some index set $\alpha \subseteq \{1, \ldots, n\}$ such that

$$\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}[\alpha, \emptyset^c]) = \operatorname{rank}(\boldsymbol{A}[\emptyset^c, \alpha]) \tag{4}$$

(that is, if there are r linearly independent rows of $\boldsymbol{A}$ such that the corresponding r columns are linearly independent), then $\boldsymbol{A}$ is rank principal; moreover, $\boldsymbol{A}[\alpha]$ is nonsingular.

4.2 The inverse of a partitioned matrix

- Given $\boldsymbol{A} \in M_n(F)$ and $\boldsymbol{A}^{-1}$ are also nonsingular. For simplicity, let $\boldsymbol{A}$ be partitioned as a 2-by-2 block matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{A}_{1,2} \\ \boldsymbol{A}_{2,1} & \boldsymbol{A}_{2,2} \end{bmatrix}$$

with $\boldsymbol{A}_{i,i} \in M_{n_i}(F)$, $i = 1, 2$, and $n_1 + n_2 = n$. A useful expression for the correspondingly **partitioned** presentation of $\boldsymbol{A}^{-1}$ is

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \left(\boldsymbol{A}_{1,1} - \boldsymbol{A}_{1,2}\boldsymbol{A}_{2,2}^{-1}\boldsymbol{A}_{2,1}\right)^{-1} & -\boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}\left(\boldsymbol{A}_{2,2} - \boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}\right)^{-1} \\ -\boldsymbol{A}_{2,2}^{-1}\boldsymbol{A}_{2,1}\left(\boldsymbol{A}_{1,1} - \boldsymbol{A}_{1,2}\boldsymbol{A}_{2,2}^{-1}\boldsymbol{A}_{2,1}\right)^{-1} & \left(\boldsymbol{A}_{2,2} - \boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}\right)^{-1} \end{bmatrix} \tag{5}$$

assuming that all the relevant inverses exist. The block diagonal terms $(\boldsymbol{A}_{1,1}^{-1}, \boldsymbol{A}_{2,2}^{-1})$ are the ***inverse of Schur complement*** with respect to $\boldsymbol{A}_{2,2}$ and $\boldsymbol{A}_{1,1}$, respectively.

- Let $\boldsymbol{A} = [a_{i,j}] \in M_n(F)$ be given and suppose that $\alpha \subset \{1, \ldots, n\}$ is an index set such that $\boldsymbol{A}[\alpha]$ is nonsingular. An important formula for $\det \boldsymbol{A}$, based on the 2-partition of $\boldsymbol{A}$ using α and α^c, is

$$\det \boldsymbol{A} = \det\left(\boldsymbol{A}[\alpha]\right)\ \det\left(\boldsymbol{A}[\alpha^c] - \boldsymbol{A}[\alpha^c, \alpha]\,\boldsymbol{A}[\alpha]^{-1}\,\boldsymbol{A}[\alpha, \alpha^c]\right) \tag{6}$$

The special matrix

$$\boldsymbol{S} := \boldsymbol{A}/\boldsymbol{A}[\alpha] = \boldsymbol{A}[\alpha^c] - \boldsymbol{A}[\alpha^c, \alpha]\,\boldsymbol{A}[\alpha]^{-1}\,\boldsymbol{A}[\alpha, \alpha^c] \tag{7}$$

is called the ***Schur complement*** of $\boldsymbol{A}[\alpha]$ in $\boldsymbol{A}$. Thus we have

$$\det \boldsymbol{A} = \det\left(\boldsymbol{A}[\alpha]\right)\ \det\left(\boldsymbol{A}/\boldsymbol{A}[\alpha]\right) \tag{8}$$

$$\begin{aligned} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{A}_{1,2} \\ \boldsymbol{A}_{2,1} & \boldsymbol{A}_{2,2} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{2,2} - \boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}/\boldsymbol{A}_{1,1} \end{bmatrix} \end{aligned} \tag{9}$$

- $\boldsymbol{A}$ is nonsingular *if and only* if both $\boldsymbol{A}_{1,1}$ and its the Schur complement $\boldsymbol{A}/\boldsymbol{A}_{1,1}$ are nonsingular, since $\det \boldsymbol{A} = \det(\boldsymbol{A}_{1,1}) \det(\boldsymbol{A}/\boldsymbol{A}_{1,1})$. If $\boldsymbol{A}$ is nonsingular, then $\det(\boldsymbol{A}/\boldsymbol{A}_{1,1}) = \det \boldsymbol{A}/\det \boldsymbol{A}_{1,1}$.

- We can have alternative expression in inverse of block matrix for Schur complement $\boldsymbol{S} := \boldsymbol{A}/\boldsymbol{A}_{1,1}$

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{A}_{1,1} + \boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}\boldsymbol{S}^{-1}\boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1} & -\boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}\boldsymbol{S}^{-1} \\ -\boldsymbol{S}^{-1}\boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1} & \boldsymbol{S}^{-1} \end{bmatrix} \tag{10}$$

4.3 The Sherman-Morrison-Woodbury formula

Suppose that a nonsingular matrix $\boldsymbol{A} = [a_{i,j}] \in M_n(F)$ has a known inverse $\boldsymbol{A}^{-1}$ and consider $\boldsymbol{B} = \boldsymbol{A} + \boldsymbol{XRY}$, in which $\boldsymbol{X}$ is n-by-r, $\boldsymbol{Y}$ is r-by-n, and $\boldsymbol{R}$ is r-by-r and nonsingular. If $\boldsymbol{B}$ and $\boldsymbol{R}^{-1} + \boldsymbol{YA}^{-1}\boldsymbol{X}$ are nonsingular, then

$$\boldsymbol{B}^{-1} = (\boldsymbol{A} + \boldsymbol{XRY})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{X}\left(\boldsymbol{R}^{-1} + \boldsymbol{YA}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{YA}^{-1} \tag{11}$$

If r is much smaller than n, then $\boldsymbol{R}$ and $\boldsymbol{R}^{-1} + \boldsymbol{YA}^{-1}\boldsymbol{X}$ may be much easier to invert than $\boldsymbol{B}$. For instance

$$\left(\boldsymbol{A} + \boldsymbol{xy}^T\right)^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{xy}^T\boldsymbol{A}^{-1}}{1 + \boldsymbol{y}^T\boldsymbol{Ax}} \tag{12}$$

if $\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{y}^T\boldsymbol{x} \neq -1$ then

$$\left(\boldsymbol{I} + \boldsymbol{xy}^T\right)^{-1} = \boldsymbol{I} - \frac{\boldsymbol{xy}^T}{1 + \boldsymbol{y}^T\boldsymbol{x}} \tag{13}$$

4.4 Complementary nullities

4.5 Rank in a partitioned matrix and rank-principal matrices

4.6 Commutativity

5 Determinant

5.1 Definition and basic properties

5.2 Elementary row and column operations

5.3 Reduced row echelon form

5.4 Compound matrices

5.5 The adjugate and the inverse

5.6 Cramers rule

5.7 Minors of the inverse

5.8 Schur complements and determinantal formulae

5.9 Determinantal identities of Sylvester and Kronecker

5.10 The Cauchy-Binet formula

Let $\boldsymbol{A} \in M_{m,k}(F)$, $\boldsymbol{B} \in M_{k,n}(F)$, and $\boldsymbol{C} = \boldsymbol{AB}$. Furthermore, let $1 \leq r \leq \min\{m, k, n\}$, and let $\alpha \subseteq \{1, \ldots, m\}$ and $\beta \subseteq \{1, \ldots, n\}$ be index sets, each of *cardinality* r. An expression for the α, β

minor of $\boldsymbol{C}$ is

$$\det\left(\boldsymbol{C}[\alpha,\beta]\right) = \sum_{\gamma} \det\left(\boldsymbol{A}[\alpha,\gamma]\right) \det\left(\boldsymbol{B}[\gamma,\beta]\right), \tag{14}$$

where the sum is taken over *all* index sets $\gamma \subseteq \{1,\ldots,k\}$ of cardinality r.

5.11 The Laplace expansion theorem

5.12 Derivative of the determinant

5.13 Adjugates and compounds

6 Equivalence relations