A Crash Course on Linear Algebra I

Ye Wang
University of North Carolina at Chapel Hill

Mathematics and Statistics For Political Research
POLI783

Vectors

- A scalar is a point on the real axis, or in $\mathbb{R}^1$.
- A P-dimensional real vector represents a point in $\mathbb{R}^P$, the P-dimensional Euclidean space.
- Conventionally, we see a vector $\mathbf{v}$ as a column and use $\mathbf{v}'$ (transpose) to denote the same vector as a row.
- But vectors can also represent functions rather than numbers.
- E.g., the gradient of a multivariate function:

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_P} \right)'$$

- Similarly, a vector of r.v.s is: $\mathbf{X} = (X_1, X_2, \ldots, X_P)'$.
- These vectors can be viewed as "points" in the corresponding function spaces.
- Such spaces share many structural similarities with the familiar Euclidean space.

Vector space

- Remember that a set is a collection of objects.
- We are often interested in sets whose elements share common operations.
- For example, any two points in $\mathbb{R}^3$ can be added and have their distance computed.
- A space is simply a set equipped with additional structure so that certain operations are defined.
- This abstraction lets us study shared properties across seemingly different kinds of objects.
- Both the set of all P-dimensional real vectors and the set of all P-dimensional random variables are vector spaces.

Vector space

- A vector space $\mathcal{V}$ is a set equipped with the following two operations:
  - addition: $\mathbf{u} + \mathbf{v} \in \mathcal{V}$,
  - scalar multiplication: $a\mathbf{v} \in \mathcal{V}$, for $a \in \mathbb{R}$.
- We require the following properties for the two operations:
  1. Closure: $\mathbf{u} + \mathbf{v} \in \mathcal{V}$, $a\mathbf{v} \in \mathcal{V}$.
  2. Commutativity: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
  3. Associativity: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
  4. Zero vector: There exists $\mathbf{0}$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
  5. Additive inverse: For every $\mathbf{v}$, there exists $-\mathbf{v}$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
  6. Distributivity: $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$ and $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$.
  7. Compatibility: $(ab)\mathbf{v} = a(b\mathbf{v})$.
  8. Identity: $1\mathbf{v} = \mathbf{v}$.
- There properties are satisfied by the set of all P-dimensional real vectors or random variables.

The dimensionality of a vector space

- Can we compute the distance between any two vectors from $\mathcal{V}$?
- Not really. We need additional structures to define a distance.
- The properties of a vector space implies that for $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_P) \in \mathcal{V}$,

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_P\mathbf{v}_P \in \mathcal{V}.$$

- We say that $\mathbf{u}$ is a linear combination of $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_P)$.
- If we can find $(c_1, c_2, \ldots, c_P)$, not all zero, such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_P\mathbf{v}_P = \mathbf{0}$, then $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_P)$ are linearly dependent.
- Otherwise, they are linearly independent.
- If there exist a set of linearly independent vectors, $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_P)$, such that any $\mathbf{u} \in \mathcal{V}$ can be expressed as their linear combination, then $\mathcal{V}$ has the dimensionality of P.
- The vectors $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_P)$ form a basis of $\mathcal{V}$.

The dimensionality of a vector space

- The dimensionality of $\mathcal{V}$ is derived from the basic properties.
- Consider the space of all Pth-degree polynomials:

$$f(x) = a_0 + a_1 x + \cdots + a_P x^P.$$

- What would be a basis of it?
- Both $(1, x, x^2, \ldots, x^P)$ and $(1, x+1, x^2+x, \ldots, x^P + x^{P-1})$ are bases.
- Therefore, bases are not unique in a vector space, but their dimensionality is fixed.
- The vector space of all integrable functions has a dimensionality of infinity.

Inner product

- We define the inner product of two real vectors, **u** and **v**, as

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{p=1}^{P} u_p v_p.$$

- We can see that $\langle \mathbf{u}, \mathbf{u} \rangle = \sum_{p=1}^{P} u_p^2 = ||\mathbf{u}||_2^2$, where $|| \cdot ||_2$ represents the vector's length.
- Inner product satisifies certain properties:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle \text{ (symmetry)},$$
$$\langle a\mathbf{u}_1 + b\mathbf{u}_2, \mathbf{v} \rangle = a\langle \mathbf{u}_1, \mathbf{v} \rangle + b\langle \mathbf{u}_2, \mathbf{v} \rangle \text{ (linearity)},$$
$$\langle \mathbf{u}, \mathbf{u} \rangle \geq 0, \langle \mathbf{u}, \mathbf{u} \rangle = 0 \iff \mathbf{u} = \mathbf{0} \text{ (positive definiteness)}.$$

- Conversely, we call all mappings from $\mathcal{V} \times \mathcal{V}$ to $\mathbb{R}$ that satisfy these properties an inner product.
- A linear space equipped with an inner product is known as an inner product space.

Inner product

- We say two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal to each other if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.
- For any inner product space $\mathcal{V}$, we can define the L^2 norm of $\mathbf{v} \in \mathcal{V}$ as $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.
- $\mathbf{v}$ is a unit vector if $||\mathbf{v}|| = 1$.
- In the vector space of all r.v.s with a finite variance, we define the inner product of X and Y as

$$\langle X, Y \rangle = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathrm{Cov}[X, Y].$$

- Then, the L^2 norm of X is $\sqrt{\langle X, X \rangle} = \sqrt{\mathrm{Var}[X]}$.
- We can show that $\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq ||\mathbf{u}||^2 ||\mathbf{v}||^2$ (the Cauchy-Schwartz inequality).
- Therefore, $\mathrm{Cov}^2[X, Y] \leq \mathrm{Var}[X]\,\mathrm{Var}[Y]$, and the correlation coefficient $|\rho| \leq 1$.
- In $\mathbb{R}^P$, it reduces to $\left(\sum_{p=1}^{P} u_p v_p\right)^2 \leq \left(\sum_{p=1}^{P} u_p^2\right)\left(\sum_{p=1}^{P} v_p^2\right)$.

Inner product

- In quantum mechanics, the position and momentum of a particle are denoted as x and p, respectively.
- Both are random variables before we measure them.
- The Cauchy-Schwartz inequality tells us that

$$\sigma_x \sigma_p \geq \mathrm{Cov}[x, p] = \frac{h}{4\pi},$$

  where h is the Planck constant.
- This is Heisenberg's uncertainty principle.
- An inner product space becomes a Hilbert space if it is complete.
- Intuitively, completeness means that whenever a sequence of points keeps getting closer and closer together, it eventually converges to a point that belongs to the same space.

The Matrix vs. a matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix}$$

Matrix

- We have seen it a few times in the semester.
- Each a_{mn} could be a function, but it is usually a real number.
- An $M \times N$ matrix $\mathbf{A}$ is a collection of N M-dimensional vectors:

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N),$$

  where $\mathbf{a}_n = (a_{1n}, a_{2n}, \ldots, a_{Mn})'$.
- Or, we see it as the collection of M N-dimensional vectors:

$$\mathbf{A} = \begin{pmatrix} \tilde{\mathbf{a}}_1' \\ \tilde{\mathbf{a}}_2' \\ \vdots \\ \tilde{\mathbf{a}}_M', \end{pmatrix} = (\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \ldots, \tilde{\mathbf{a}}_M)'.$$

  where $\tilde{\mathbf{a}}_m' = (a_{m1}, a_{m2}, \ldots, a_{mN})'$.
- A real vector with length P can be seen as a $P \times 1$ matrix.

Matrix

- $\mathbf{A}$ is a square matrix if $M = N$.
- A square matrix is a diagnol matrix if all of its off-diagnoal elements are zero.
- We denote it as $\mathbf{A} = diag\{a_1, a_2, \ldots, a_N\}$.
- An identity matrix is $\mathbf{I} = diag\{1, 1, \ldots, 1\}$.
- A square matrix is symmetric if $a_{mn} = a_{nm}$ for any $1 \leq m \leq M$ and $1 \leq n \leq N$.
- A square matrix is called upper (lower) triangular if all of its entries below (above) the main diagonal are zero, i.e., $a_{ij} = 0$ whenever $i > j$ $(i < j)$.

Matrix operation: rank, trace, and transpose

- The rank of a matrix is the largest number of linearly independent row (or column) vectors in the matrix.
- It measures the dimensionality of the space spanned by the row or column vectors.
- A matrix is full rank if $rank(\mathbf{A}) = \min\{M, N\}$.
- For a square matrix, its trace is the sum of its diagnol elements: $tr(\mathbf{A}) = \sum_{i=1}^{N} a_{ii}$.
- The transpose of an $M \times N$ matrix $\mathbf{A}$ is a $N \times M$ matrix denoted as $\mathbf{A}'$:

$$\mathbf{A}' = \begin{pmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_N' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{M1} \\ a_{12} & a_{22} & \dots & a_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1N} & a_{2N} & \dots & a_{MN} \end{pmatrix}.$$

- We can verify that 1) $(\mathbf{A}')' = \mathbf{A}$, and 2) $\mathbf{A}' = \mathbf{A}$ for symmetric matrices.

Matrix operation: addition

- For two $M \times N$ matrices, $\mathbf{A}$ and $\mathbf{B}$, their sum is defined as

$$\mathbf{A}+\mathbf{B} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1N}+b_{1N} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2N}+b_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}+b_{M1} & a_{M2}+b_{M2} & \dots & a_{MN}+b_{MN} \end{pmatrix}$$
$$=(\mathbf{a}_1+\mathbf{b}_1, \mathbf{a}_2+\mathbf{b}_2, \dots, \mathbf{a}_N+\mathbf{b}_N).$$

- Addition is well-defined only for matrices with the same dimensionality.
- $\mathbf{A}+\mathbf{B} = \mathbf{B}+\mathbf{A}$ (commutative law).
- $(\mathbf{A}+\mathbf{B})' = \mathbf{A}'+\mathbf{B}'$.

Matrix operation: muliplication

- The product of two matrices is defined by the inner product of vectors they consist of.
- Suppose $\mathbf{A}$ is an $M \times N$ matrix: $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_M)'$, and $\mathbf{B}$ is a $N \times P$ matrix: $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_P)$, then

$$\mathbf{AB} = \begin{pmatrix} \sum_{k=1}^N a_{1k} b_{1k} & \sum_{k=1}^N a_{1k} b_{2k} & \ldots & \sum_{k=1}^N a_{1k} b_{Pk} \\ \sum_{k=1}^N a_{2k} b_{1k} & \sum_{k=1}^N a_{2k} b_{2k} & \ldots & \sum_{k=1}^N a_{2k} b_{Pk} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^N a_{Mk} b_{1k} & \sum_{k=1}^N a_{Mk} b_{2k} & \ldots & \sum_{k=1}^N a_{Mk} b_{Pk} \end{pmatrix}$$
$$= \begin{pmatrix} \langle \mathbf{a}_1, \mathbf{b}_1 \rangle & \langle \mathbf{a}_1, \mathbf{b}_2 \rangle & \ldots & \langle \mathbf{a}_1, \mathbf{b}_P \rangle \\ \langle \mathbf{a}_2, \mathbf{b}_1 \rangle & \langle \mathbf{a}_2, \mathbf{b}_2 \rangle & \ldots & \langle \mathbf{a}_2, \mathbf{b}_P \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{a}_M, \mathbf{b}_1 \rangle & \langle \mathbf{a}_M, \mathbf{b}_2 \rangle & \ldots & \langle \mathbf{a}_M, \mathbf{b}_P \rangle \end{pmatrix}.$$

- $\mathbf{AB} \neq \mathbf{BA}$ in general, and their dimensionalities may differ.
- $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ and $\mathbf{I}_M\mathbf{A} = \mathbf{A}\mathbf{I}_N = \mathbf{A}$.
- $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, $(\mathbf{A}'\mathbf{A})' = \mathbf{A}'\mathbf{A}$.

Matrix operation: muliplication

- In particular, for two N-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$,

$$\mathbf{x}'\mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k=1}^{N} x_k y_k.$$

- Meanwhile, $\mathbf{x}\mathbf{y}'$ is a $N \times N$ matrix.
- Then, we can see that

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} \sum_{k=1}^{N} a_{1k} x_k \\ \sum_{k=1}^{N} a_{2k} x_k \\ \vdots \\ \sum_{k=1}^{N} a_{Mk} x_k \end{pmatrix} = \begin{pmatrix} \langle \mathbf{a}_1, \mathbf{x} \rangle \\ \langle \mathbf{a}_2, \mathbf{x} \rangle \\ \vdots \\ \langle \mathbf{a}_M, \mathbf{x} \rangle \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1'\mathbf{x} \\ \mathbf{a}_2'\mathbf{x} \\ \vdots \\ \mathbf{a}_M'\mathbf{x} \end{pmatrix}.$$

- $\mathbf{I}\mathbf{x} = \mathbf{x}$, $\mathbf{A}\mathbf{0} = \mathbf{0}$, and $(\mathbf{A}\mathbf{x})' = \mathbf{x}'\mathbf{A}'$.

Matrix operation: muliplication

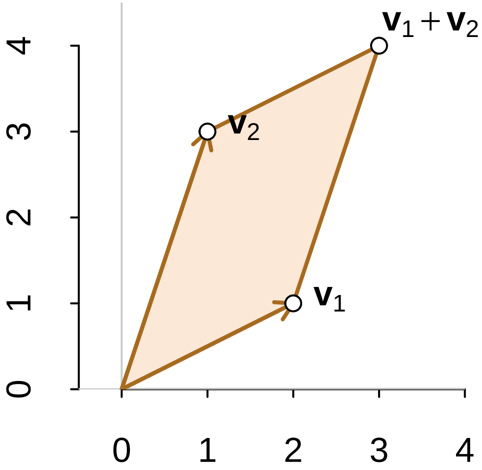
- Consider the following two matrices:

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ -1 & 3 \\ 4 & 0 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 2 & 0 \\ -2 & 1 & 5 \end{pmatrix}.$$

- Can we compute **AB**? If so, what is the matrix's dimensionality?
- How about **BA**?
- If $\mathbf{x} = (1, 0, 1)'$, what is $\mathbf{Ax}$?

Matrix operation: determinant

- For a full-rank square matrix $\mathbf{A}$, its column (row) vectors comprise a parallelepiped (high-dimensional parallelogram) in space:



Matrix operation: determinant

- The volume of this parallelepiped is known as the determinant of the matrix, denoted as $det(\mathbf{A})$.
- The determinant is directional (just like the integral).
- For a 2×2 matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $det(\mathbf{A}) = ad - bc$.
- For higher-order matrices, the computation is complex thus we skip it.
- $det(\mathbf{A}) \neq 0$ if and only if $\mathbf{A}$ is full rank.
- Otherwise, the parallelepiped is not well-defined.
- For square matrices $\mathbf{A}$ and $\mathbf{B}$, $det(\mathbf{AB}) = det(\mathbf{A})det(\mathbf{B})$.
- Therefore, $det\left(\mathbf{A}^{-1}\right) = 1/det(\mathbf{A})$.
- For a diagonal matrix $\mathbf{A} = diag\{a_1, a_2, \ldots, a_N\}$, $det(\mathbf{A}) = \prod_{i=1}^{N} a_i$.

Matrix operation: inverse

- We say $\mathbf{A}^{-1}$ is the inverse of a square matrix $\mathbf{A}$ if $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.
- A matrix is invertible if and only if it is full rank if and only if $det(\mathbf{A}) \neq 0$.
- We say $\mathbf{A}$ is an orthogonal matrix if $\mathbf{A}^{-1} = \mathbf{A}'$.
- Each column (row) vector in an orthogonal matrix has a length of 1.
- Any two column (row) vectors in an orthogonal matrix are orthogonal to each other.
- The determinant of an orthogonal matrix is 1.
- A system of linear equations can be expressed as $\mathbf{Ax} = \mathbf{b}$, thus its solution is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.
- If $\mathbf{A}$ is not invertible, the equations may still have a solution, but the solution won't be unique.

Matrices as mappings

- Each matrix can be seen as a linear transformation or mapping.
- Suppose $\mathbf{A}$ is a full-rank $M \times N$ matrix and $\mathbf{x}$ is a N-dimensional vector.
- $\mathbf{y} = \mathbf{A}\mathbf{x}$ transforms $\mathbf{x}$ to an M-dimensional vector.
- The mapping's domain is $\mathbb{R}^N$ and its range is $\left\{\mathbf{y} \in \mathbb{R}^M : \mathbf{y} = \mathbf{A}\mathbf{x}\right\}$.
- The range is a linear space, and we call it the image space of $\mathbf{A}$.
- The dimensionality of the image space is exactly $\mathbf{A}$'s rank.
- The linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ have solutions if and only if $\mathbf{b}$ belongs to $\mathbf{A}$'s image space.
- We define the kernel space of $\mathbf{A}$ as the collection of $\mathbf{x} \in \mathbb{R}^N$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$.
- The system of linear equations has a unique solution if and only if the kernel space only contains $\mathbf{0}$.

Matrices as mappings

- We can show that N equals the kernel space's dimensionality plus the image space's dimensionality.
- The system of linear equations has a unique solution if $rank(\mathbf{A}) = N$.
- If so, $\mathbf{A}$ is a full-rank square matrix, and the number of independent equations equals the number of unknowns.
- It also means that $det(\mathbf{A}) \neq 0$.
- Note that the system $\mathbf{Ax} = \mathbf{0}$ always has a solution of $\mathbf{0}$.
- It has non-zero solutions when $det(\mathbf{A}) = 0$.
- We can simplify any vector by representing it as a linear combination of the basis vectors.
- We can similarly simplify any linear mapping via its decomposition.