CS 660: Mathematical Foundations of Analytics

Dr. Francis Parisi

Pace University

Spring 2018

Sequences

Definition

A sequence is a set of numbers $a_1, a_2, a_3, \ldots$ in a correspondence with the natural numbers $\mathbb{N}$ and formed according to a definite rule

- Each number in the sequence is called a term; a_n is called the n^{th} term
- A sequence is either finite or infinite
- The sequence $a_1, a_2, a_3, \ldots$ is also designated by $\{a_n\}$

Example

The set of numbers $1, 1/3, 1/5, 1/7, \ldots$ is an infinite sequence whose n^{th} term is $a_n = \frac{1}{2n-1}$, where $n = 1, 2, 3, \ldots$

Sequences

- A number l is called the limit of an infinite sequence $a_1, a_2, a_3, \ldots$ if for any positive number ϵ we can find a positive number N depending on ϵ such that $a_n - l < \epsilon$ for all integers $n > N$. In such case we write $\lim_{n \to \infty} a_n = l$
- In other words, if a sequence has a limit l its terms get closer and closer to l
- If the limit of a sequence exists, the sequence is called *convergent*; otherwise, it is called *divergent*
- If a limit exists, it is unique

Sequences

- If $a_n \leq M$ for $n = 1, 2, 3, \ldots$, where M is a constant (independent of n), we say that the sequence $\{a_n\}$ is bounded above and M is called an upper bound
- If $a_n \geq m$, the sequence is bounded below and m is called a lower bound If $m \leq a_n \leq M$ the sequence is called bounded and is often indicated by $a_n \leq P$
- Every convergent sequence is bounded, but the converse is not necessarily true
- If $a_{n+1} \geq a_n$ the sequence is called monotonic increasing; if $a_{n+1} > a_n$ it is called strictly increasing
- Similarly, if $a_{n+1} \leq a_n$ the sequence is called monotonic decreasing, while if $a_{n+1} < a_n$ it is strictly decreasing

Sequences

Example

1. The sequence $1, 1.1, 1.11, 1.111, \ldots$ is bounded and monotonic increasing. It is also strictly increasing.

2. The sequence $1, -1, 1, -1, 1, \ldots$ is bounded but not monotonic increasing or decreasing.

3. The sequence $-1, -1.5, -2, -2.5, -3, \ldots$ is monotonic decreasing and not bounded. However, it is bounded above.

Sequences

- $\underline{M}$ is called the least upper bound (lub) of the sequence $\{a_n\}$ if $a_n \leq \underline{M}$, $n = 1, 2, 3, \ldots$ while at least one term is greater than $\underline{M} - \epsilon$ for any $\epsilon > 0$
- $\overline{m}$ is called the greatest lower bound (glb) of the sequence $\{a_n\}$ if $a_n \geq \overline{m}$, $n = 1, 2, 3, \ldots$ while at least one term is greater than $\overline{m} + \epsilon$ for any $\epsilon > 0$
- The lub of a sequence is also called the **supremum**
- The glb of a sequence is also called the **infimum**

- A number $\bar{l}$ is called the *limit supremum*, ($\limsup$ or $\overline{\lim}$) of the sequence $\{a_n\}$ if infinitely many terms of the sequence are greater than $\bar{l} - \epsilon$ while only a finite number of terms are greater than $\bar{l} + \epsilon$, where ϵ is any positive number
- A number $\underline{l}$ is called the *limit infimum*, ($\liminf$ or $\underline{\lim}$) of the sequence $\{a_n\}$ if infinitely many terms of the sequence are less than $\bar{l} + \epsilon$ while only a finite number of terms are less than $\bar{l} - \epsilon$, where ϵ is any positive number
- These correspond to least and greatest limiting points of general sets of numbers

Sequences

Definition

We say a sequence $\{a_n\}$ is *Cauchy*, if for any $\epsilon > 0$, there is an N such that $|a_n - a_m| < \epsilon$, if $n, m \geq N$.

As the following theorem states, the Cauchy property is a necessary and sufficient condition for the convergence of a sequence

Theorem

Let $\{a_n\}$ be a sequence. Then $\{a_n\}$ converges if and only if $\{a_n\}$ is Cauchy.

Sequences

Infinite Series

Let $\{an\}$ be a sequence of real numbers

The corresponding infinite series is $\sum_{i=1}^{n} a_n$

The sequence of partial sums $\{Sn\}$ is given by

$$S_n = \sum_{i=1}^{n} a_i$$

If $S_n \to S$, as $n \to \infty$, then we say the series $\sum_{i=1}^{n} a_i$ converges to S and write S as the infinite series,

$$S = \sum_{i=1}^{n} a_i \tag{2.1}$$

Assume that the series $\sum_{i=1}^{n} a_n$ converges to S. Then, using (2.1), for any $\epsilon > 0$, there exists an N such that

$$\left| \sum_{n=N+1}^{\infty} a_n \right| = \left| S - \sum_{j=1}^{n} a_j \right| < \epsilon \tag{2.2}$$

In other words, the tail of a convergent series becomes arbitrarily small as n gets large

Sequences

Partial sums can be obtained easily in some cases like for the geometric series

Example (Geometric Series)

For $a \in \mathbb{R}$ consider the sequence of partial sums given by

$$S_n = \sum_{i=0}^{n} a^i$$

We can write this as

$$(1-a)S_n = 1 - a^{n+1}$$

so that

$$S_n = \frac{1 - a^{n+1}}{1 - a}$$

Sequences

Example (Geometric Series Cont'd...)

If $|a| < 1$ then $a^{n+1} \to 0$ as $n \to \infty$ and the geometric series converges to $(1-a)^{-1}$ or

$$\sum_{n=0}^{\infty} a^n = \frac{1}{1-a} \tag{2.3}$$

A series $\sum a_n$ converges absolutely if $\sum |a_n|$ converges

Absolute convergence implies convergence

Derivatives

Let $f(x)$ be a real valued function defined on an interval (a,b) of real numbers, including $(-\infty,\infty)$

f is differentiable at x with derivative $f'(x)$ if the following limit exists

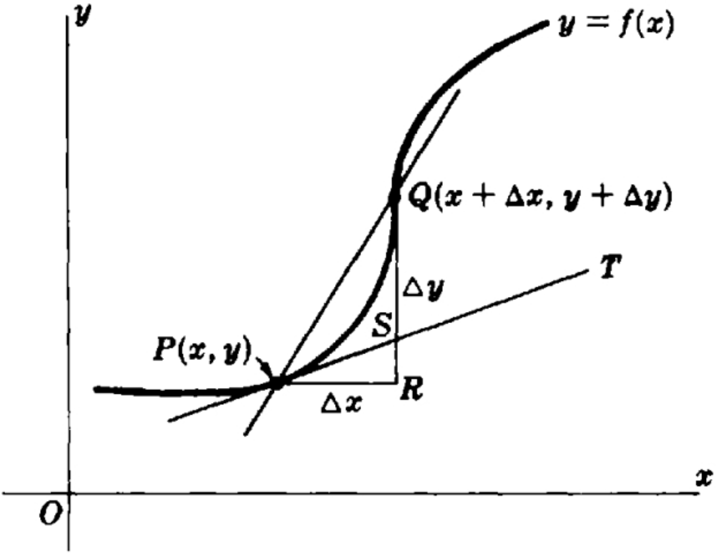
$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} = f'(x)$$

We often write $f'(x) = \frac{d}{dx} f(x)$

If f is differentiable then f is continuous

Derivatives

Graphical Representation



Derivatives

In the previous graph, The line connecting points Q and P is the secant line

$\dfrac{\Delta y}{\Delta x}$ is the *average rate of change* and the slope of the secant line

We get the *instantaneous rate of change* when calculate the derivative

$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} = f'(x)$$

As Q approaches P ie, $\Delta x \to 0$, the secant line approaches the tangent line and $\dfrac{dy}{dx}$ is the slope of the tangent line

Derivatives

Example

EXAMPLE 2: Find the derivative of $y = f(x) = x^2 + 3x$ with respect to x at $x = x_0$. Use this to find the value of the derivative at (a) $x_0 = 2$ and (b) $x_0 = -4$.

$$\begin{aligned}
y_0 &= f(x_0) = x_0^2 + 3x_0 \\
y_0 + \Delta y &= f(x_0 + \Delta x) = (x_0 + \Delta x)^2 + 3(x_0 + \Delta x) \\
&= x_0^2 + 2x_0\,\Delta x + (\Delta x)^2 + 3x_0 + 3\,\Delta x \\
\Delta y &= f(x_0 + \Delta x) - f(x_0) = 2x_0\,\Delta x + 3\,\Delta x + (\Delta x)^2 \\
\frac{\Delta y}{\Delta x} &= \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = 2x_0 + 3 + \Delta x
\end{aligned}$$

The derivative at $x = x_0$ is

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} (2x_0 + 3 + \Delta x) = 2x_0 + 3$$

(a) At $x_0 = 2$, the value of the derivative is $2(2) + 3 = 7$.
(b) At $x_0 = -4$, the value of the derivative is $2(-4) + 3 = -5$.

Derivatives

Exercise

Find $\frac{dy}{dx}$, given $y = x^3 - x^2 - 4$ using the definition of the derivative.

Derivatives

The *right-hand* derivative of $f(x)$ at $x = x_0$ is defined as

$$f'_+(x_0) = \lim_{\Delta x \to 0+} \frac{f(x_0 + \Delta x) - f(x)}{\Delta x}$$

if the limit exists – Δx approaches 0 from above

The *left-hand* derivative of $f(x)$ at $x = x_0$ is defined as

$$f'_-(x_0) = \lim_{\Delta x \to 0-} \frac{f(x_0 + \Delta x) - f(x)}{\Delta x}$$

if the limit exists – Δ x approaches 0 from below

A function f has a derivative at $x = x_0$ if and only if $f'_+(x_0) = f'_-(x_0)$

Derivatives

Some useful formulæ for differentiation

$$\frac{d}{dx}x = 1; \frac{d}{dx}x^r = rx^{r-1} \text{ (not differentialble at } 0 \text{ for } r < 1)$$

$$\frac{d}{dx}\left[a\,f(x) + b\;g(x)\right] = a\,f'(x) + b\;g'(x)$$

$$\frac{d}{dx}\left[f(x)g(x)\right] = f'(x)g(x) + f(x)g'(x)$$

$$\frac{d}{dx}\left[f(x)/g(x)\right] = \left[f'(x)g(x) + f(x)g'(x)\right]/g(x)^2$$

$$\frac{d}{dx}f(x)^a = a\,f'(x)^{a-1}$$

$$\frac{d}{dx}f\left[g(x)\right] = f'\left[g(x)\right]g'(x)$$

$$\frac{d}{dx}e^{f(x)} = f'(x)e^{f(x)}$$

Derivatives

If a function $f(x)$ is strictly increasing or decreasing onan interval (a,b), its inverse function $f^{-1}(y)$ exists on (a,b) and its derivative is

$$\frac{d}{dy}f^{-1}(y) = \frac{1}{f'\left[f^{-1}(y)\right]}$$

Derivatives

Theorem (Mean Value Theorem)

Let $f(x)$ be continuous n the interval $[a, b]$ and differentiable on (a, b). Then there is a point xi such that $a < \xi < b$ and

$$f(b) = f(a) + (b - a)f'(\xi) \tag{2.4}$$

Derivatives

Mean Value Theorem

If $f(x)$ is differentiable in an open neighborhood of x_0 and $f'(x)$ is continuous at x_0 then for some ξ between x and x_0 we can write (2.4) as

$$\begin{aligned} f(x) &= f(x_0) + (x - x_0)f'(\xi) \\ &= f(x_0) + (x - x_0)f'(x_0) + (x - x_0)\left[f'(\xi) - f'(x_0)\right] \\ &= f(x_0) + (x - x_0)f'(x_0) + o(x - x_0) \end{aligned} \tag{2.5}$$

the little o notation is defined by

$$a = o(b) \text{ iff } \frac{a}{b} \to 0 \text{ as } b \to 0$$

Derivatives

Given a function $f(x)$ with derivative $f'(x)$, if $f'(x)$ is differentiable then

$$\frac{d}{dx}f'(x) = \frac{d^2}{dx^2}f(x) = f''(x) = f^{(2)}(x)$$

is the second derivative of $f(x)$

In general, the n^{th} derivative $f^{(n)}(x)$ is found by repeated differentiation

Derivatives

Partial Derivatives

In the case of a function of two or more variables we introduce the concept of **partial derivatives**

These derivatives are taken with respect to one variable while the other variables are held fixed

For the *partial* derivative with respect to x, we use the notation $\frac{\partial}{\partial x}$

Derivatives

Partial Derivatives

Example

Let $f(x, y) = 2x^2 \exp\{-x^2 - y^2\}$. Then

$$\frac{\partial}{\partial x} f(x, y) = 4x \exp\{-x^2 - y^2\} - 4x^3 \exp\{-x^2 - y^2\}$$
$$\frac{\partial}{\partial y} f(x, y) = -4x^2 y \exp\{-x^2 - y^2\}$$

Higher order partial derivatives are found from repeated partial differentiation, where the variables may be mixed, ie.,

$$\frac{\partial^2}{\partial xy}, \ \frac{\partial^2}{\partial y^2}, \ \frac{\partial^2}{\partial x^2}, \ \frac{\partial^2}{\partial yx}$$

Integrals

We motivate our discussion of integrals by considering the classic geometric approach and Riemann sums

- Let $F(x)$ be a function defined on ,$[a, b]$, and we partition $[a, b]$ into n subintervals

$$[a + (i - 1)\Delta x, a + i\Delta x]$$

  where $i = 1, 2, 3, \ldots, n$ and $\Delta x = \frac{b-a}{n}$
- Let ξ_i represent the midpoint of the i^{th} interval, then the Riemann sum is

$$R_n = R_n(f, a, b, \Delta) = \sum_{i=1}^{n} f(\xi_i)\Delta x$$

  which is an average

Integrals

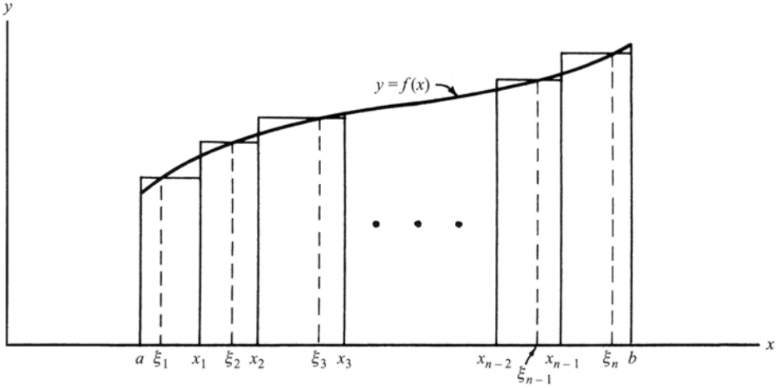
If $f(x)$ is continuous then R_n converges to a limit as $n \to \infty$ and is denoted

$$\lim_{n\to\infty} R_n = \int_a^b f(x)dx$$

and is known as the **definite interval** of $f(x)$ over $[a, b]$

If $f(x) \geq 0$ on the interval then R_n approximates the area under the curve $y = f(x)$ between a and b

Integrals



Integrals

Oftentimes the bounds may be infinite as occurs frequently in probability theory, $\int_a^\infty f(x)dx$ where $f(x)$ is a density function, so the integral is given by the limit

$$\lim_{h \to \infty} \int_a^h f(x)dx = \int_a^\infty f(x)dx$$

Integrals

Theorem (Fundamental Theorem of Calculus)

Let the function $F(x)$ be differentiable on the interval $[a, b]$ with derivative $f(x)$. If $f(x)$ is continuous on $[a, b]$ then

$$F(b) - F(a) = \int_a^b f(x)dx$$

.

Example

Since x^n is the derivative of $x^{n+1}/(n+1)$ we have

$$\int_a^b x^n dx = \left.\frac{x^{n+1}}{n+1}\right|_a^b = \frac{1}{n+1}\left[b^{n+1} - a^{n+1}\right]$$

Integrals

Some properties of integrals

$$\int_a^b [c\,f(x) + d\;g(x)]\,dx = c\int_a^b f(x)dx + \;d\int_a^b g(x)dx \tag{2.6}$$

$$a < c < b \Rightarrow \int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx \tag{2.7}$$

$$\int_a^b f(x)dx = -\int_b^a f(x)dx \tag{2.8}$$

$$f(x) \geq g(x) \text{ on } [a,b] \Rightarrow \int_a^b f(x)dx \geq \int_a^b g(x)dx \tag{2.9}$$

Integrals

Change-of-Variable Technique

- Suppose $g(x)$ is differentiable on $[a, b]$ and that $F(x)$ is differentiable on the range of $g(x)$
- From the chain rule we have

$$D_x F\left[g(x)\right] = F'\left[g(x)\right] g'(x)$$

- We get

$$\int_a^b F'\left[g(x)\right] g'(x) = F\left[g(x)\right]\Big|_a^b = F\left[g(b)\right] - F\left[g(a)\right] = \int_{g(a)}^{g(b)} F'(u)du$$

Integrals

Change-of-Variable Technique

To simplify the notation we typically write $u = g(x)$ and $du = g'(x)dx$

Exercise

Use the change-of variable technique to find the following

$$\int_2^3 x \exp\left\{-x^2\right\}.$$

What is u? What is du? Evaluate the integral.

Integrals

Integration by parts

Suppose $u(x)$ and $v(x)$ are differentiable funcitons on $[a, b]$ whose second derivatives exist

The product rule for differentiation gives

$$\frac{d}{dx}\left[u(x)v(x)\right] = u'(x)v(x) + u(x)v'(x)$$

If we solve for $u(x)v'(x)$ and integrate we get

$$\int_a^b u(x)v'(x) = u(x)v(x)\Big|_a^b - \int_a^b u'(x)v(x)d(x)$$

This is **Integration by Parts**

Integrals

Exercise

1. Evaluate

$$\int (x+2)\sin(x^2+4x-6)dx$$

2. Determine

$$\int_e^{e^2} \frac{dx}{x(\ln x)^3}$$

3. Evaluate

$$\int 3^{\sqrt{2x+1}}dx$$

Multiple Integration

- We can extend the concept of integration to n-dimensions
- Consider the two dimensional case

Let $f(x,y)$ be a continuous funciton of two variables, x and y, defined on a bounded rectangle A

We partition A into mn sub-rectangles and sum over all the sub-rectangles

$$R_{m,n} = R_{m,n}(f, A, \Delta x, \Delta y) = \sum_{i=1}^{m} \sum_{i=1}^{n} f(x_i, y_i) \Delta x_i \Delta y_i$$

where (x_i, y_i) is an interior point

Multiple Integration

If the largest sub-rectangle converges to 0 as $m, n \to \infty$ then $R_{m,n}$ has a limit which is the double integral of f over A

$$\iint\limits_{A} f(x,y)dxdy$$

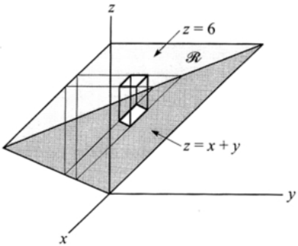
Multiple Integrals

Multiple integrals can be computed as *iterated* integrals
Integrate with respect to one variable then integrate the result wrt the other variable

Exercise

Find the volume of the region bounded by

$$z = x + y, z = 6, x = 0, y = 0, z = 0$$



Matrix Algebra

A matrix $\boldsymbol{A}$ over a field K or, simply, a matrix $\boldsymbol{A}$ (when K is implicit) is a rectangular array of numbers usually presented in the following form:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Matrix Algebra

- Matrices are denoted by uppercase bold letters $\boldsymbol{A}$
- A **matrix** $\boldsymbol{A}$ having m rows and n columns is an $m \times n$ matrix
- To emphasize the *size* of the matrix we can write $\boldsymbol{A} = [m \times n]$, for a matrix with m rows and n columns
- $\boldsymbol{A}$ is a **square matrix** if $m = n$
- We specify an element of $\boldsymbol{A}$ by a_{ij}, $i = 1, \ldots, m$, and $j = 1, \ldots, n$, and $\boldsymbol{A} = (a_{ij})$
- The inverse of a (nonsingular) matrix is denoted $\boldsymbol{A}^{-1}$
- The transpose of a matrix is denoted $\boldsymbol{A}^T$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \qquad \boldsymbol{A}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Matrix Algebra

Let and $\boldsymbol{A} = [a_{ij}]$ and $\boldsymbol{B} = [b_{ij}]$ be two matrices with the same size, say $m \times n$ matrices. The sum of $\boldsymbol{A}$ and $\boldsymbol{B}$, written $\boldsymbol{A} + \boldsymbol{B}$, is the matrix obtained by adding corresponding elements from $\boldsymbol{A}$ and $\boldsymbol{B}$. That is,

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{11} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$$

Matrix Algebra

The product of the matrix $\boldsymbol{A}$ by a scalar k, written $k \times \boldsymbol{A}$ or simply $k\boldsymbol{A}$, is the matrix obtained by multiplying each element of $\boldsymbol{A}$ by k. That is,

$$k\boldsymbol{A} = \begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{bmatrix}$$

Matrix Algebra

Example (Matrix Addition)

$$\begin{bmatrix} -2 & 1 & 0 \\ 3 & -4 & 2 \end{bmatrix} + \begin{bmatrix} 5 & -4 & 3 \\ -1 & 2 & 6 \end{bmatrix} = \begin{bmatrix} (-2)+5 & 1+(-4) & 0+3 \\ 3+(-1) & (-4)+2 & 2+6 \end{bmatrix} = \begin{bmatrix} 3 & -3 & 3 \\ 2 & -2 & 8 \end{bmatrix}$$

A **row matrix** has a single row and a **column matrix** has a single column

Matrix Algebra

Definition

Let $\boldsymbol{A}$ be a row matrix and $\boldsymbol{B}$ be a column matrix, each with n entries. Hence $\boldsymbol{A}$ is a $1 \times n$ matrix and $\boldsymbol{B}$ is an $n \times 1$ matrix. The **inner product** or **dot product** $\boldsymbol{A} \cdot \boldsymbol{B}$ of $\boldsymbol{A}$ and $\boldsymbol{B}$ is obtained by adding the products of the corresponding elements of $\boldsymbol{A}$ and $\boldsymbol{B}$. That is, if

$$\boldsymbol{A} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \text{ and } \boldsymbol{B} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

then

$$\boldsymbol{A} \cdot \boldsymbol{B} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

Matrix Algebra

The product $\boldsymbol{AB}$ of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined if the number of columns of $\boldsymbol{A}$ equals the number of rows of $\boldsymbol{B}$

Definition

Let $\boldsymbol{A}$ be an $m \times n$ matrix and $\boldsymbol{B}$ an $n \times p$ matrix, where $\boldsymbol{A} = [a_{ij}]$ with $1 \le i \le m$ and $1 \le j \le n$, and $\boldsymbol{B} = [b_{jk}]$, with $1 \le j \le n$ and $1 \le k \le p$. Then the **product** $\boldsymbol{C} = \boldsymbol{AB}$ of $\boldsymbol{A}$ and $\boldsymbol{B}$ is the $m \times p$ matrix, where $\boldsymbol{C} = [c_{ik}]$ where $1 \le i \le m$ and $1 \le k \le p$ such that

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}.$$

To calculate the (i,k)-entry c_{ik} of $\boldsymbol{C}$, we compute the inner product of row i of $\boldsymbol{A}$ (which has n columns) and column k of $\boldsymbol{B}$ (which has n rows).

Matrix Algebra

Definition

For an $n \times n$ square matrix $\boldsymbol{A}$, the **powers** of $\boldsymbol{A}$ are defined by

$$\boldsymbol{A}^1 = \boldsymbol{A} \text{ and } \boldsymbol{A}^k = \underbrace{\boldsymbol{A}\boldsymbol{A}\cdots\boldsymbol{A}}_{k \text{ factors}} \text{ for each integer } k \geq 2.$$

For example, if $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$, then

$$\boldsymbol{A}^1 = \boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}, \ \boldsymbol{A}^2 = \boldsymbol{A}\boldsymbol{A} = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}, \text{ and}$$

$$\boldsymbol{A}^3 = \boldsymbol{A}\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^2 = \begin{bmatrix} 5 & 6 \\ 3 & 2 \end{bmatrix}.$$

Matrix Algebra

- The identity matrix is denoted by $\boldsymbol{I}$ or $\boldsymbol{I_n}$ to emphasize the dimension

$$\boldsymbol{I_3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- The $n \times n$ matrix of ones is denoted $\boldsymbol{U_n}$

$$\boldsymbol{U_3} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Matrix Algebra

- All vectors are column vectors and are denoted by lowercase bold letters, $\boldsymbol{a}$
- There are exceptions
  - $\boldsymbol{J}$ denotes a column vector of ones
  - and $\mathbf{0}$ denotes either a column vector or matrix of zeros
- We may write $\boldsymbol{J}_m,\ \mathbf{0}_m,$ or $\mathbf{0}_{m,n}$ to emphasize the dimensions
- Transposing a column vector gives a row vector, $\boldsymbol{a}^T$

Matrix Algebra

- In general, we may partition a matrix $\boldsymbol{A}$ into a set of rectangular matrices, provided they are of appropriate dimensions

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

Example

$$A = \left[\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right]$$

Matrix Algebra

Definition

The **determinant** of a matrix A is a scaling factor of the transformation the matrix

$$det(\boldsymbol{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

For a 3×3 matrix we have

$$det(\boldsymbol{A}) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

Matrix Algebra

Definition

A matrix $\boldsymbol{A}$ is singular if its determinant is zero. A singular matrix does not have an inverse.

Definition

Let $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}$ be vectors representing the columns of matrix $\boldsymbol{A}$. Then the $\boldsymbol{u_i}$ are linearly dependent if there exist scalars k_i not all zeros such that

$$k_1\boldsymbol{u_1} + k_2\boldsymbol{u_2} + \cdots + k_n\boldsymbol{u_n} = 0$$

Matrix Algebra

- The **rank** of a matrix $\boldsymbol{A}$ is the number of linearly independent rows or columns and is denoted by $r(\boldsymbol{A})$
- A **square matrix** is a matrix with the same number of rows and columns
- The **trace** of a square matrix is the sum of its diagonals denoted by $tr(\mathrm{mod}A)$

$$tr(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}$$

- For to matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$

Matrix Algebra

- If $\boldsymbol{A}$ is a square matrix of size n, the scalar, λ, is called an eigenvalue of $\boldsymbol{A}$ if it is a root of the polynomial defined by

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0$$

  where $|\cdot|$ is the determinant of a matrix

- Associated with each eigenvalue is and eigenvector $\boldsymbol{p}$ determined by the equation

$$\boldsymbol{A}\boldsymbol{p} = \lambda \boldsymbol{p}$$

Matrix Algebra

The trace, rank, and determinant of $\boldsymbol{A}$ can be found by its eigenvalues:

1. $tr(\boldsymbol{A}) = \sum_i \lambda_i$
2. $r(\boldsymbol{A}) =$ the number of nonzero eigenvalues
3. $|\boldsymbol{A}| = \prod_i \lambda_i$

Matrix Algebra

The function

$$q(\boldsymbol{y}) = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y} = \sum_i \sum_j a_{ij} y_i y_j$$

is called a quadratic form defined by the matrix $\boldsymbol{A}$

- A quadratic form is **positive definite** if $q(\boldsymbol{y}) > 0$ for all nonzero vectors $\boldsymbol{y}$
- If $q(\boldsymbol{y}) \geq 0$ and $q(\boldsymbol{y}) = 0$ for some nonzero $\boldsymbol{y}$ then $q(\boldsymbol{y})$ is **positive semi-definite**

Matrix Algebra

A square matrix is said to be **orthogonal** if and only if $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}}$. Orthogonal matrices have the following properties

1. The rows and columns of $\boldsymbol{P}$ are orthogonal and have length one, so
$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{I}$$
2. $|\boldsymbol{P}| = \pm 1$
3. $-1 \leq p_{ii} \leq 1$

The matrix of eigenvectors of a symmetric matrix is orthogonal

Matrix Algebra

An $n \times n$ matrix $\boldsymbol{A}$ is **idempotent** if $\boldsymbol{AA} = \boldsymbol{A}$. Idempotent matrices have the following properties

1. $r(\boldsymbol{A}) = n$ implies that $\boldsymbol{A} = \boldsymbol{I}$
2. The nonzero eigenvalues of $\boldsymbol{A}$ are equal to one
3. The trace of $\boldsymbol{A}$ equals the rank
4. If $\boldsymbol{A}$ is symmetric and idempotent then $\boldsymbol{A}$ is at least positive semi-definite
5. Suppose that the matrices $\boldsymbol{A_i}$, $i = 1, 2, \ldots, m$ are symmetric

$$\boldsymbol{A} = \sum \boldsymbol{A_i}, \; r(\boldsymbol{A_i}) = r_i, \text{ and } r(\boldsymbol{A}) = r$$

Let $\boldsymbol{A}$ be any n-square matrix. Then $\boldsymbol{A}$ can be represented by (or is similar to) a diagonal matrix $D = \operatorname{diag}(k_1, k_2, \ldots, k_n)$ if and only if there exists a basis S consisting of (column) vectors $u_1, u_2 \ldots u_n$ such that

$$\begin{aligned}
\boldsymbol{A}u_1 &= k_1 u_1 \\
\boldsymbol{A}u_2 &= \qquad\quad k_2 u_2 \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\boldsymbol{A}u_n &= \qquad\qquad\qquad\quad k_n u_n
\end{aligned}$$

In such a case, $\boldsymbol{A}$ is said to be diagonalizable. Furthermore, $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$, where $\boldsymbol{P}$ is the nonsingular matrix whose columns are, respectively, the basis vectors $u_1, u_2, \ldots, u_n$.

Matrix Algebra

Theorem

An n-square matrix $\boldsymbol{A}$ is similar to a diagonal matrix $\boldsymbol{D}$ if and only if $\boldsymbol{A}$ has n linearly independent eigenvectors. In this case, the diagonal elements of $\boldsymbol{D}$ are the corresponding eigenvalues and $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$, where $\boldsymbol{P}$ is the matrix whose columns are the eigenvectors.

Matrix Algebra

Example

Let

$$\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} \text{ and let } \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \text{ and } \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Then

$$\boldsymbol{A}v_1 = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix} = v_1$$

and

$$\boldsymbol{A}v_2 = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} = 4v_2$$

Matrix Algebra

Example (Continued...)

Thus, $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are eigenvectors of $\boldsymbol{A}$ belonging, respectively, to the eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 4$.

v_1 and v_2 are linearly independent and form a basis of $\mathbb{R}^2$.

Accordingly, $\boldsymbol{A}$ is diagonalizable. Furthermore, let $\boldsymbol{P}$ be the matrix whose columns are the eigenvectors $v1$ and $v2$. That is, let

$$\boldsymbol{P} = \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix} \text{ and } \boldsymbol{P}^{-1} = \begin{bmatrix} 1/3 & -1/3 \\ 2/3 & 1/3 \end{bmatrix}$$

Matrix Algebra

Example (Continued...)

Then A is similar to the diagonal matrix

$$\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} 1/3 & -1/3 \\ 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

As expected, the diagonal elements 1 and 4 in $\boldsymbol{D}$ are the eigenvalues corresponding, respectively, to the eigenvectors v_1 and v_2, which are the columns of $\boldsymbol{P}$. Thus $\boldsymbol{A}$ has the factorization

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} = \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1/3 & -1/3 \\ 2/3 & 1/3 \end{bmatrix}$$

An algorithm for computing eigenvalues and eigenvectors for a given square matrix $\boldsymbol{A}$ and for determining whether or not a nonsingular matrix $\boldsymbol{P}$ exists such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is diagonal.

First we must define the *characteristic polynomial*

Definition

Let $\boldsymbol{A} = [a_{ij}]$ be an n-square matrix. The matrix $\boldsymbol{M} = \boldsymbol{A} - t\boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is the n-square identity matrix and t is an indeterminate, may be obtained by subtracting t down the diagonal of $\boldsymbol{A}$. The negative of $\boldsymbol{M}$ is the matrix $t\boldsymbol{I}_n - \boldsymbol{A}$, and its determinant $\Delta(t) = \det(t\boldsymbol{I}_n - \boldsymbol{A}) = (-1)^n \det(\boldsymbol{A} - t\boldsymbol{I}_n)$ which is a polynomial in t of degree n and is called the *characteristic polynomial* of $\boldsymbol{A}$.

For a characteristic polynomial of degree 2 the equation becomes

$$\Delta(t) = t^2 - tr(\boldsymbol{A}) + \det(\boldsymbol{A})$$

Matrix Algebra

Diagonalization Algorithm

Step1. Find the characteristic polynomial $\Delta(t)$ of $\boldsymbol{A}$.

Step2. Find the roots of $\Delta(t)$ to obtain the eigenvalues of $\boldsymbol{A}$.

Step3. For each eigenvalue λ of $\boldsymbol{A}$, repeat (a) and (b) below

- (a) Form the matrix $\boldsymbol{M} = \boldsymbol{A} - \lambda \boldsymbol{I}$ by subtracting λ down the diagonal of $\boldsymbol{A}$.
- (b) Find a basis for the solution space of the homogeneous system $\boldsymbol{MX} = 0$. (These basis vectors are linearly independent eigenvectors of $\boldsymbol{A}$ belonging to λ.)

Step4. Consider the collection S = $\{v_1, v_2, \ldots, v_m\}$ of all eigenvectors obtained in Step 3.

- (a) If $m \neq n$, then $\boldsymbol{A}$ is not diagonalizable.
- (b) If $m = n$, then $\boldsymbol{A}$ is diagonalizable. Specifically, let $\boldsymbol{P}$ be the matrix whose columns are the eigenvectors $v_1, v_2, \ldots, v_n$. Then

$$D = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

  where λ_i is the eigenvalue corresponding to the eigenvector v_i.

Exercise

Apply the diagonalization algorithm to

$$\boldsymbol{A} = \begin{bmatrix} 4 & 2 \\ 3 & -1 \end{bmatrix}$$

to find the eigenvalues, eigenvectors, and a nonsingular matrix $\boldsymbol{P}$ to diagonalize $\boldsymbol{A}$.

References

Hocking, R. R. (1996).
Methods and Applications of Linear Models.
John Wiley & Sons, Inc., New York.

Hogg, R. V., McKean, J. W. and Craig, A. T. (2018).
Introduction to Mathematical Statistics.
Pearson, Boston, eighth edition.

Kabacoff, R. I. (2015).
R in Action.
Manning, Shelter Island, NY, second edition.

Lipschutz, S. and Lipson, M. L. (2009).
Linear Algebra.
Schaum's Outline. McGraw-Hill, New York, fourth edition.

Wrede, R. and Spiegel, M. R. (2010).
Advanced Calculus.
Schaum's Outline. McGraw-Hill, New York, third edition.

References (cont.)

Wu, J. and Coggeshall, S. (2012).
Foundations of Predictive Analytics.
CRC Press, Boca Raton, FL.