COMP 761: Lecture 14 – Calculus II

David Rolnick

October 5, 2020

Problem

Let X, A be $n \times n$ matrices, with X the matrix with all entries equal to t^2. What is the derivative $\frac{d}{dt}\operatorname{tr}(XA)$?

(Please don't post your ideas in the chat just yet, we'll discuss the problem soon in class.)

Course Announcements

Course Announcements

- Problem 3 clarification: “prove it is possible to determine the remainder” is the same as “prove that there is exactly one possible value for this remainder.”

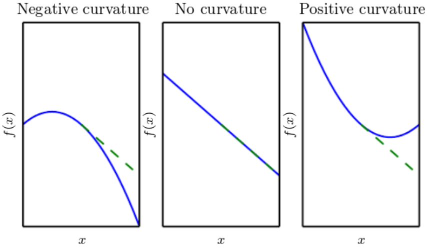
Course Announcements

- Problem 3 clarification: “prove it is possible to determine the remainder” is the same as “prove that there is exactly one possible value for this remainder.”
- Office hours right after class today.

Double derivatives

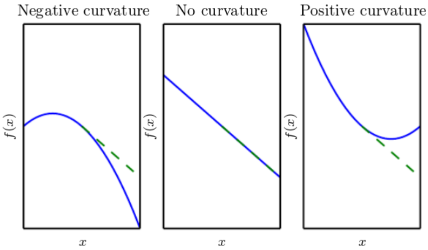
Double derivatives

- Graphically, measures the *curvature*:



Double derivatives

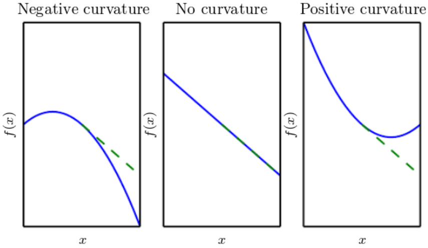
- Graphically, measures the *curvature*:



- If a function $f(x)$ has $f''(x) \geq 0$ for all x, we say it is *convex*.

Double derivatives

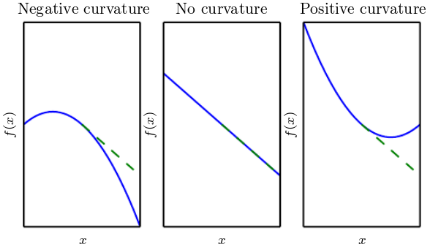
- Graphically, measures the *curvature*:



- If a function $f(x)$ has $f''(x) \geq 0$ for all x, we say it is *convex*.
- If a function $f(x)$ has $f''(x) \leq 0$ for all x, we say it is *concave*.

Double derivatives

- Graphically, measures the *curvature*:



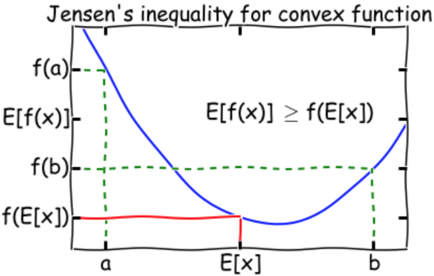
- If a function $f(x)$ has $f''(x) \geq 0$ for all x, we say it is *convex*.
- If a function $f(x)$ has $f''(x) \leq 0$ for all x, we say it is *concave*.
- **Most functions are neither convex nor concave.**

Jensen's inequality

Jensen's inequality

- *Jensen's inequality*: if $f(x)$ is a convex function, then

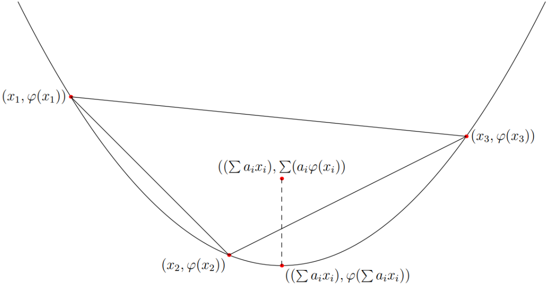
$$\frac{f(x_1) + \ldots + f(x_n)}{n} \geq f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$



Jensen's inequality

- *Jensen's inequality*: if $f(x)$ is a convex function, then

$$\frac{f(x_1) + \ldots + f(x_n)}{n} \geq f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$



Jensen's inequality

Jensen's inequality

- If $f(x)$ is concave, then $-f(x)$ is convex, so

$$\frac{-f(x_1) - \ldots - f(x_n)}{n} \geq -f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$

therefore

$$\frac{f(x_1) + \ldots + f(x_n)}{n} \leq f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$

Jensen's inequality

- If $f(x)$ is concave, then $-f(x)$ is convex, so

$$\frac{-f(x_1) - \ldots - f(x_n)}{n} \geq -f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$

therefore

$$\frac{f(x_1) + \ldots + f(x_n)}{n} \leq f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$

Jensen's inequality

- If $f(x)$ is concave, then $-f(x)$ is convex, so

$$\frac{-f(x_1) - \ldots - f(x_n)}{n} \geq -f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$

therefore

$$\frac{f(x_1) + \ldots + f(x_n)}{n} \leq f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$

- What if $f''(x) = 0$? (both convex and concave)

Jensen's inequality

- If $f(x)$ is concave, then $-f(x)$ is convex, so

$$\frac{-f(x_1) - \ldots - f(x_n)}{n} \geq -f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$

therefore

$$\frac{f(x_1) + \ldots + f(x_n)}{n} \leq f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$

- What if $f''(x) = 0$? (both convex and concave)
- Then, $f(x) = ax + b$ (linear function) and:

$$\frac{f(x_1) + \ldots + f(x_n)}{n} = f\left(\frac{x_1 + \ldots + x_n}{n}\right)$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \quad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

- Can we use Jensen's inequality? How?

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \quad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

- Can we use Jensen's inequality? How?
- When you see addition of things and multiplication of those same things, you should think about log, since $\log(xy) = \log(x) + \log(y)$.

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \quad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

- Can we use Jensen's inequality? How?
- When you see addition of things and multiplication of those same things, you should think about log, since $\log(xy) = \log(x) + \log(y)$.
- Is $\log(x)$ concave, convex, or neither?

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

- Can we use Jensen's inequality? How?
- When you see addition of things and multiplication of those same things, you should think about log, since $\log(xy) = \log(x) + \log(y)$.
- Is $\log(x)$ concave, convex, or neither?
- We have

$$\frac{d^2}{dx^2} \log(x) = \frac{d}{dx} \frac{1}{x}$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

- Can we use Jensen's inequality? How?
- When you see addition of things and multiplication of those same things, you should think about log, since $\log(xy) = \log(x) + \log(y)$.
- Is $\log(x)$ concave, convex, or neither?
- We have

$$\frac{d^2}{dx^2} \log(x) = \frac{d}{dx} \frac{1}{x} = \frac{d}{dx} x^{-1}$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

- Can we use Jensen's inequality? How?
- When you see addition of things and multiplication of those same things, you should think about log, since $\log(xy) = \log(x) + \log(y)$.
- Is $\log(x)$ concave, convex, or neither?
- We have

$$\frac{d^2}{dx^2}\log(x) = \frac{d}{dx}\frac{1}{x} = \frac{d}{dx}x^{-1} = (-1)x^{-2}$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

- Can we use Jensen's inequality? How?
- When you see addition of things and multiplication of those same things, you should think about log, since $\log(xy) = \log(x) + \log(y)$.
- Is $\log(x)$ concave, convex, or neither?
- We have

$$\frac{d^2}{dx^2}\log(x) = \frac{d}{dx}\frac{1}{x} = \frac{d}{dx}x^{-1} = (-1)x^{-2} = \frac{-1}{x^2}$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

- Can we use Jensen's inequality? How?
- When you see addition of things and multiplication of those same things, you should think about log, since $\log(xy) = \log(x) + \log(y)$.
- Is $\log(x)$ concave, convex, or neither?
- We have

$$\frac{d^2}{dx^2} \log(x) = \frac{d}{dx} \frac{1}{x} = \frac{d}{dx} x^{-1} = (-1)x^{-2} = \frac{-1}{x^2} < 0.$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

- Can we use Jensen's inequality? How?
- When you see addition of things and multiplication of those same things, you should think about log, since $\log(xy) = \log(x) + \log(y)$.
- Is $\log(x)$ concave, convex, or neither?
- We have

$$\frac{d^2}{dx^2}\log(x) = \frac{d}{dx}\frac{1}{x} = \frac{d}{dx}x^{-1} = (-1)x^{-2} = \frac{-1}{x^2} < 0.$$

- So $\log(x)$ is concave.

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n \geq 0$$

- Can we use Jensen's inequality? How?
- When you see addition of things and multiplication of those same things, you should think about log, since $\log(xy) = \log(x) + \log(y)$.
- Is $\log(x)$ concave, convex, or neither?
- We have

$$\frac{d^2}{dx^2} \log(x) = \frac{d}{dx} \frac{1}{x} = \frac{d}{dx} x^{-1} = (-1)x^{-2} = \frac{-1}{x^2} < 0.$$

- So $\log(x)$ is concave.
- That means we could apply Jensen's inequality to $-\log(x)$.

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \quad \text{for } a_1, a_2, \ldots, a_n > 0$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n > 0$$

- Jensen's inequality applied to $-\log(x)$ is:

$$\frac{-\log(a_1) - \log(a_2) - \cdots - \log(a_n)}{n} \geq -\log\left(\frac{a_1 + a_2 + \cdots + a_n}{n}\right).$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n > 0$$

- Jensen's inequality applied to $-\log(x)$ is:

$$\frac{-\log(a_1) - \log(a_2) - \cdots - \log(a_n)}{n} \geq -\log\left(\frac{a_1 + a_2 + \cdots + a_n}{n}\right).$$

- Equivalently:

$$\log\left(\frac{a_1 + \cdots + a_n}{n}\right) \geq \frac{\log(a_1) + \cdots + \log(a_n)}{n}$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n > 0$$

- Jensen's inequality applied to $-\log(x)$ is:

$$\frac{-\log(a_1) - \log(a_2) - \cdots - \log(a_n)}{n} \geq -\log\left(\frac{a_1 + a_2 + \cdots + a_n}{n}\right).$$

- Equivalently:

$$\log\left(\frac{a_1 + \cdots + a_n}{n}\right) \geq \frac{\log(a_1) + \cdots + \log(a_n)}{n} = \frac{\log(a_1 \cdots a_n)}{n}$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n > 0$$

- Jensen's inequality applied to $-\log(x)$ is:

$$\frac{-\log(a_1) - \log(a_2) - \cdots - \log(a_n)}{n} \geq -\log\left(\frac{a_1 + a_2 + \cdots + a_n}{n}\right).$$

- Equivalently:

$$\log\left(\frac{a_1 + \cdots + a_n}{n}\right) \geq \frac{\log(a_1) + \cdots + \log(a_n)}{n} = \frac{\log(a_1 \cdots a_n)}{n}$$
$$= \log\left((a_1 \cdots a_n)^{1/n}\right).$$

Problem

Prove the AM-GM Inequality:

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}, \qquad \text{for } a_1, a_2, \ldots, a_n > 0$$

- Jensen's inequality applied to $-\log(x)$ is:

$$\frac{-\log(a_1) - \log(a_2) - \cdots - \log(a_n)}{n} \geq -\log\left(\frac{a_1 + a_2 + \cdots + a_n}{n}\right).$$

- Equivalently:

$$\log\left(\frac{a_1 + \cdots + a_n}{n}\right) \geq \frac{\log(a_1) + \cdots + \log(a_n)}{n} = \frac{\log(a_1 \cdots a_n)}{n}$$
$$= \log\left((a_1 \cdots a_n)^{1/n}\right).$$

- Taking exponentials of both sides gives us AM-GM.

Maxima/minima

Maxima/minima

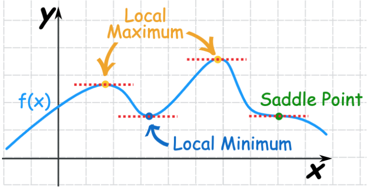
- *Critical points* of a differentiable function $f(x)$ are x with $f'(x) = 0$.

Maxima/minima

- *Critical points* of a differentiable function $f(x)$ are x with $f'(x) = 0$.
- If a point x is a *local max* or a *local min* of $f(x)$, then $f'(x) = 0$.

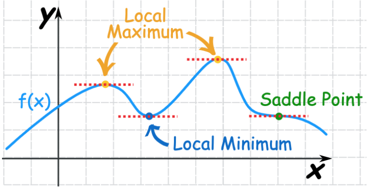
Maxima/minima

- *Critical points* of a differentiable function $f(x)$ are x with $f'(x) = 0$.
- If a point x is a *local max* or a *local min* of $f(x)$, then $f'(x) = 0$.
- Pictorially, this makes sense:



Maxima/minima

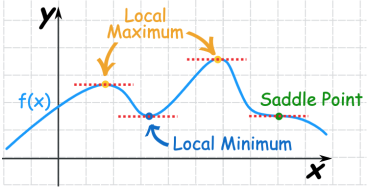
- *Critical points* of a differentiable function $f(x)$ are x with $f'(x) = 0$.
- If a point x is a *local max* or a *local min* of $f(x)$, then $f'(x) = 0$.
- Pictorially, this makes sense:



- Algebraically, also makes sense – consider a local maximum x: If the slope $f'(x)$ were positive then increasing x would make $f(x)$ even higher, while if the slope $f'(x)$ were negative then making x smaller would increase $f(x)$.

Maxima/minima

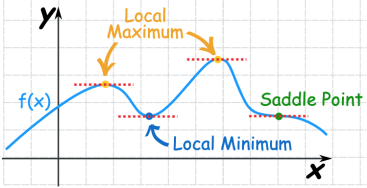
- *Critical points* of a differentiable function $f(x)$ are x with $f'(x) = 0$.
- If a point x is a *local max* or a *local min* of $f(x)$, then $f'(x) = 0$.
- Pictorially, this makes sense:



- Algebraically, also makes sense – consider a local maximum x: If the slope $f'(x)$ were positive then increasing x would make $f(x)$ even higher, while if the slope $f'(x)$ were negative then making x smaller would increase $f(x)$.
- So how to determine if local maximum or minimum?

Maxima/minima

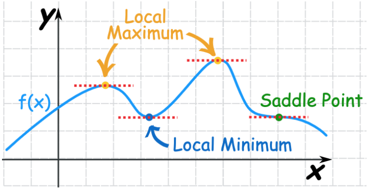
- *Critical points* of a differentiable function $f(x)$ are x with $f'(x) = 0$.
- If a point x is a *local max* or a *local min* of $f(x)$, then $f'(x) = 0$.
- Pictorially, this makes sense:



- Algebraically, also makes sense – consider a local maximum x: If the slope $f'(x)$ were positive then increasing x would make $f(x)$ even higher, while if the slope $f'(x)$ were negative then making x smaller would increase $f(x)$.
- So how to determine if local maximum or minimum?
- If $f'(x) = 0$ and $f''(x) > 0$, then local maximum.

Maxima/minima

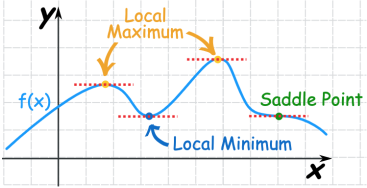
- *Critical points* of a differentiable function $f(x)$ are x with $f'(x) = 0$.
- If a point x is a *local max* or a *local min* of $f(x)$, then $f'(x) = 0$.
- Pictorially, this makes sense:



- Algebraically, also makes sense – consider a local maximum x: If the slope $f'(x)$ were positive then increasing x would make $f(x)$ even higher, while if the slope $f'(x)$ were negative then making x smaller would increase $f(x)$.
- So how to determine if local maximum or minimum?
- If $f'(x) = 0$ and $f''(x) > 0$, then local maximum.
- If $f'(x) = 0$ and $f''(x) < 0$, then local minimum.

Maxima/minima

- *Critical points* of a differentiable function $f(x)$ are x with $f'(x) = 0$.
- If a point x is a *local max* or a *local min* of $f(x)$, then $f'(x) = 0$.
- Pictorially, this makes sense:



- Algebraically, also makes sense – consider a local maximum x: If the slope $f'(x)$ were positive then increasing x would make $f(x)$ even higher, while if the slope $f'(x)$ were negative then making x smaller would increase $f(x)$.
- So how to determine if local maximum or minimum?
- If $f'(x) = 0$ and $f''(x) > 0$, then local maximum.
- If $f'(x) = 0$ and $f''(x) < 0$, then local minimum.
- If $f'(x) = 0$ and $f''(x) = 0$, then neither (saddle point).

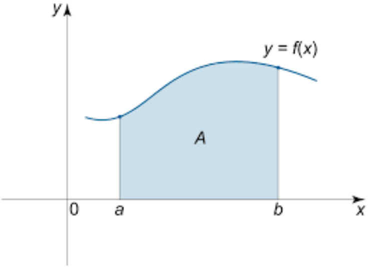
The integral

The integral

- The integral

$$\int_a^b f(x)dx$$

is the area under the curve of the function $f(x)$ between $x = a$ and $x = b$.



Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus states that

$$\frac{d}{dx}\int_a^x f(t)dt = f(x).$$

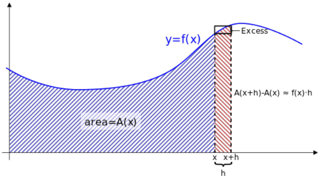
Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus states that

$$\frac{d}{dx}\int_a^x f(t)dt = f(x).$$

- Writing $A(x)$ as the area under the curve from a to x:

$$\frac{d}{dx}A(x) = \lim_{h\to 0}\frac{A(x+h)-A(x)}{h}.$$



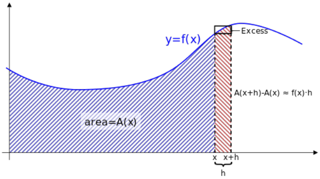
Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus states that

$$\frac{d}{dx}\int_a^x f(t)dt = f(x).$$

- Writing $A(x)$ as the area under the curve from a to x:

$$\frac{d}{dx}A(x) = \lim_{h\to 0}\frac{A(x+h)-A(x)}{h}.$$



- Can see that for $h \approx 0$, we have $A(x+h) \approx A(x) + hf(x)$, so

$$\frac{d}{dx}A(x) = \lim_{h\to 0}\frac{hf(x)}{h} = f(x).$$

The integral

Another form of the Fundamental Theorem of Calculus:

$$\int_a^b g'(t)dt = g(b) - g(a).$$

The integral

Another form of the Fundamental Theorem of Calculus:

$$\int_a^b g'(t)dt = g(b) - g(a).$$

- Proved similarly.

The integral

Another form of the Fundamental Theorem of Calculus:

$$\int_a^b g'(t)dt = g(b) - g(a).$$

- Proved similarly.
- So to compute the area under a curve $f(x)$ we just to find a function $g(x)$ such that $g'(x) = f(x)$.

The integral

Another form of the Fundamental Theorem of Calculus:

$$\int_a^b g'(t)dt = g(b) - g(a).$$

- Proved similarly.
- So to compute the area under a curve $f(x)$ we just to find a function $g(x)$ such that $g'(x) = f(x)$.
- Example, area under parabola t^2 from -1 to 1.

$$\int_{-1}^{1} t^2 dt.$$

The integral

Another form of the Fundamental Theorem of Calculus:

$$\int_a^b g'(t)dt = g(b) - g(a).$$

- Proved similarly.
- So to compute the area under a curve $f(x)$ we just to find a function $g(x)$ such that $g'(x) = f(x)$.
- Example, area under parabola t^2 from -1 to 1.

$$\int_{-1}^{1} t^2 dt.$$

- We have $\frac{d}{dt}\frac{1}{3}t^3 = t^2$, so

$$\int_{-1}^{1} t^2 dt = \frac{1}{3}(1)^3 - \frac{1}{3}(-1)^3 = \frac{2}{3}.$$

Partial derivatives

Partial derivatives

- We can also compute the *partial derivative* of a function in many variables with respect to one of those variables.

Partial derivatives

- We can also compute the *partial derivative* of a function in many variables with respect to one of those variables.
- Essentially just hold all the other variables constant.

Partial derivatives

- We can also compute the *partial derivative* of a function in many variables with respect to one of those variables.
- Essentially just hold all the other variables constant.
- Written e.g. $\partial/\partial x$ instead of d/dx.

Partial derivatives

- We can also compute the *partial derivative* of a function in many variables with respect to one of those variables.
- Essentially just hold all the other variables constant.
- Written e.g. $\partial/\partial x$ instead of d/dx.
- For example:

$$\frac{\partial}{\partial x}(x^2 + 2y^2 + xy) = 2x + y$$
$$\frac{\partial}{\partial y}(x^2 + 2y^2 + xy) = 4y + x.$$

The gradient

The gradient

- The *gradient* ∇f of a multivariable function $f(x_1, \dots, x_n)$ is the vector of partial derivatives with respect to the variables:

$$\nabla f = \left[\begin{array}{ccc} {}^{\partial f}\!/\!_{\partial x_1} & \cdots & {}^{\partial f}\!/\!_{\partial x_n} \end{array} \right].$$

The gradient

- The *gradient* ∇f of a multivariable function $f(x_1, \ldots, x_n)$ is the vector of partial derivatives with respect to the variables:

$$\nabla f = \left[\begin{array}{ccc} {}^{\partial f}\!/_{\partial x_1} & \cdots & {}^{\partial f}\!/_{\partial x_n} \end{array} \right].$$

- For example:

$$\nabla(x^2 + y^2 + 2z^2) = \left[\begin{array}{ccc} 2x & 2y & 4z \end{array} \right].$$

The gradient

The gradient

- Can use gradient to estimate the amount that f changes.

The gradient

- Can use gradient to estimate the amount that f changes.
- In one variable, we have:

$$f(x + \epsilon) \approx f(x) + \epsilon f'(x).$$

The gradient

- Can use gradient to estimate the amount that f changes.
- In one variable, we have:

$$f(x+\epsilon) \approx f(x) + \epsilon f'(x).$$

- In 2 variables what is $f(x+\epsilon_x, y+\epsilon_y)$?

The gradient

- Can use gradient to estimate the amount that f changes.
- In one variable, we have:

$$f(x+\epsilon) \approx f(x) + \epsilon f'(x).$$

- In 2 variables what is $f(x+\epsilon_x, y+\epsilon_y)$?
- Well, we have:

$$f(x+\epsilon_x, y) \approx f(x,y) + \epsilon_x \frac{\partial}{\partial x} f(x,y).$$

The gradient

- Can use gradient to estimate the amount that f changes.
- In one variable, we have:

$$f(x+\epsilon) \approx f(x) + \epsilon f'(x).$$

- In 2 variables what is $f(x+\epsilon_x, y+\epsilon_y)$?
- Well, we have:

$$f(x+\epsilon_x, y) \approx f(x,y) + \epsilon_x \frac{\partial}{\partial x} f(x,y).$$

so

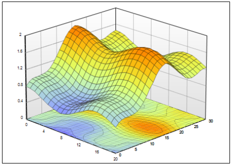
$$\begin{aligned}
f(x+\epsilon_x, y+\epsilon_y) &\approx f(x+\epsilon_x, y) + \epsilon_y \frac{\partial}{\partial y} f(x+\epsilon_x, y) \\
&\approx \left(f(x,y) + \epsilon_x \frac{\partial}{\partial x} f(x,y) \right) + \epsilon_y \frac{\partial}{\partial y} f(x+\epsilon_x, y) \\
&\approx f(x,y) + \epsilon_x \frac{\partial}{\partial x} f(x,y) + \epsilon_y \frac{\partial}{\partial y} f(x,y) \\
&= f(x,y) + \nabla f \cdot [\epsilon_x, \epsilon_y].
\end{aligned}$$

The gradient

The gradient

- In general:

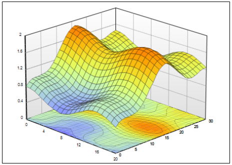
$$f(x_1 + \epsilon_1, \ldots, x_n + \epsilon_n) \approx f(x_1, \ldots, x_n) + (\nabla f) \cdot \begin{bmatrix} \epsilon_1 & \cdots & \epsilon_n \end{bmatrix}.$$



The gradient

- In general:

$$f(x_1+\epsilon_1,\ldots,x_n+\epsilon_n) \approx f(x_1,\ldots,x_n) + (\nabla f)\cdot \begin{bmatrix} \epsilon_1 & \cdots & \epsilon_n \end{bmatrix}.$$

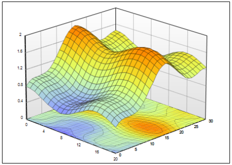


- This is why gradient descent/ascent works.

The gradient

- In general:

$$f(x_1+\epsilon_1,\dots,x_n+\epsilon_n) \approx f(x_1,\dots,x_n) + (\nabla f)\cdot\left[\begin{array}{ccc}\epsilon_1 & \cdots & \epsilon_n\end{array}\right].$$



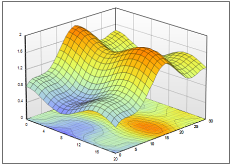
- This is why gradient descent/ascent works.
- If $\left[\begin{array}{ccc}\epsilon_1 & \cdots & \epsilon_n\end{array}\right]$ has fixed length, which direction maximizes the difference:

$$f(x_1+\epsilon_1,\dots,x_n+\epsilon_n) - f(x_1,\dots,x_n)?$$

The gradient

- In general:

$$f(x_1+\epsilon_1,\ldots,x_n+\epsilon_n) \approx f(x_1,\ldots,x_n) + (\nabla f)\cdot\begin{bmatrix}\epsilon_1 & \cdots & \epsilon_n\end{bmatrix}.$$



- This is why gradient descent/ascent works.
- If $\begin{bmatrix}\epsilon_1 & \cdots & \epsilon_n\end{bmatrix}$ has fixed length, which direction maximizes the difference:

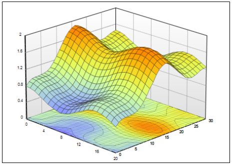
$$f(x_1+\epsilon_1,\ldots,x_n+\epsilon_n) - f(x_1,\ldots,x_n)?$$

- Dot product maximized when vectors aligned, so $\begin{bmatrix}\epsilon_1 & \cdots & \epsilon_n\end{bmatrix}$ should point along gradient (∇f).

The gradient

- In general:

$$f(x_1+\epsilon_1,\dots,x_n+\epsilon_n) \approx f(x_1,\dots,x_n) + (\nabla f)\cdot\begin{bmatrix}\epsilon_1 & \cdots & \epsilon_n\end{bmatrix}.$$



- This is why gradient descent/ascent works.
- If $\begin{bmatrix}\epsilon_1 & \cdots & \epsilon_n\end{bmatrix}$ has fixed length, which direction maximizes the difference:

$$f(x_1+\epsilon_1,\dots,x_n+\epsilon_n) - f(x_1,\dots,x_n)?$$

- Dot product maximized when vectors aligned, so $\begin{bmatrix}\epsilon_1 & \cdots & \epsilon_n\end{bmatrix}$ should point along gradient (∇f).
- Likewise, greatest *decrease* when $\begin{bmatrix}\epsilon_1 & \cdots & \epsilon_n\end{bmatrix}$ pointing along negative gradient $(-\nabla f)$.

Multivariate chain rule

Multivariate chain rule

- Multivariate function $f(x_1, \ldots, x_n)$, where $x_1 = x_1(t), \ldots, x_n = x_n(t)$.

Multivariate chain rule

- Multivariate function $f(x_1, \ldots, x_n)$, where $x_1 = x_1(t), \ldots, x_n = x_n(t)$.
- What is the derivative of f with respect to t?

Multivariate chain rule

- Multivariate function $f(x_1, \ldots, x_n)$, where $x_1 = x_1(t), \ldots, x_n = x_n(t)$.
- What is the derivative of f with respect to t?
- If we move t a small amount ϵ, then $x_1, \ldots, x_n$ change by:

$$\frac{d}{dt}x_1(t)\epsilon, \ldots, \frac{d}{dt}x_n(t)\epsilon.$$

Multivariate chain rule

- Multivariate function $f(x_1, \ldots, x_n)$, where $x_1 = x_1(t), \ldots, x_n = x_n(t)$.
- What is the derivative of f with respect to t?
- If we move t a small amount ϵ, then $x_1, \ldots, x_n$ change by:

$$\frac{d}{dt}x_1(t)\epsilon, \ldots, \frac{d}{dt}x_n(t)\epsilon.$$

- Then, the change in f is (by the gradient formula):

$$(\nabla f) \cdot \begin{bmatrix} \frac{d}{dt}x_1(t)\epsilon & \cdots & \frac{d}{dt}x_n(t)\epsilon \end{bmatrix}.$$

Multivariate chain rule

- Multivariate function $f(x_1, \ldots, x_n)$, where $x_1 = x_1(t), \ldots, x_n = x_n(t)$.
- What is the derivative of f with respect to t?
- If we move t a small amount ϵ, then $x_1, \ldots, x_n$ change by:

$$\frac{d}{dt}x_1(t)\epsilon, \ldots, \frac{d}{dt}x_n(t)\epsilon.$$

- Then, the change in f is (by the gradient formula):

$$(\nabla f) \cdot \begin{bmatrix} \frac{d}{dt}x_1(t)\epsilon & \cdots & \frac{d}{dt}x_n(t)\epsilon \end{bmatrix}.$$

- So have the *(multivariate) chain rule*:

$$\begin{aligned} \frac{d}{dt}f &= (\nabla f) \cdot \begin{bmatrix} \frac{d}{dt}x_1(t) & \cdots & \frac{d}{dt}x_n(t) \end{bmatrix} \\ &= \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}. \end{aligned}$$

Chain rule with partial derivatives

- So have the *(multivariate) chain rule*:

$$\frac{d}{dt}f = (\nabla f) \cdot \begin{bmatrix} \frac{d}{dt}x_1(t) & \cdots & \frac{d}{dt}x_n(t) \end{bmatrix}$$
$$= \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}.$$

Chain rule with partial derivatives

- So have the *(multivariate) chain rule*:

$$\frac{d}{dt}f = (\nabla f) \cdot \begin{bmatrix} \frac{d}{dt}x_1(t) & \cdots & \frac{d}{dt}x_n(t) \end{bmatrix}$$
$$= \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}.$$

- Example: if $x = t^2$ and $y = e^t$, then

$$\frac{d}{dt}(x^2 + y^2) = (2x)(2t) + (2y)(e^t)$$
$$= (2t^2)(2t) + (2e^t)(e^t)$$
$$= 4t^3 + 2e^{2t}.$$

Next time!

Calculus III