Mathematics for Computer Science B: Analysis

LECTURE 2

2024/25

SOPHIE.STEVENS@BRISTOL.AC.UK

Workshop 1 summary:

5. In 1965, Moore's law came about: Moore[1] predicted that

> The complexity for minimum component costs has increased at a rate of roughly a factor of two per year. Certainly over the short term this rate can be expected to continue, if not to increase.

We're going write this as a differential equation.

(a) Identify the variable that is changing.

(b) In words, what is your function describing?

(c) What differential equation captures Moore's law? You're looking to write something of the form $\frac{d}{dx}f(x) = g(x)$ (with your choice of variable names, and with $g(x)$ capturing Moore's law).

$$\frac{d}{dt}C(t) = C(t-1)$$

Workshop 1 summary:

- $\log(x) = \ln(x)$ for Extra Question 1
- Show that $\frac{\mathrm{d}}{\mathrm{dx}} \log(\mathrm{x}) = \frac{1}{\mathrm{x}}$
- $e^{\ln(x)} = \ln(e^x) = x$
- Use the chain rule to differentiate both sides

Today we’ll look at

Maxima and minima

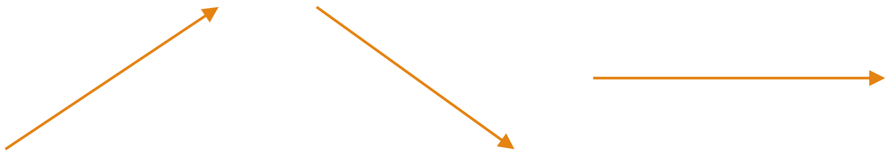
Multivariate differentiation

Gradient operator

Hessian matrix

Maxima and minima

- A gradient is either increasing or decreasing or constant



- Unless it’s constant everywhere, if the gradient is constant then it’s just finished increasing or decreasing. The graph could (locally) look like:

- Local maximum Local minimum

Maxima and minima

- A graph has a local maximum at c if there exists ϵ so that $\mathrm{f}(c) \geq f(x)$ for $|x - c| \leq \epsilon$
- A graph has a local minimum at c if there exists ϵ so that $\mathrm{f}(c) \leq f(x)$ for $|x - c| \leq \epsilon$

- A graph has a global/absolute maximum at c if $\mathrm{f}(c) \geq f(x)$ for all x in the domain
- A graph has a global/absolute minimum at c if $\mathrm{f}(c) \leq f(x)$ for all x in the domain

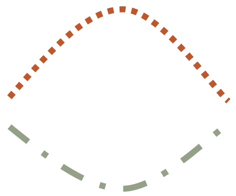
- Extreme Value Theorem: if f is continuous on a closed interval $[\mathrm{a}, \mathrm{b}]$ then f attains an absolute maximum value $f(c)$ and an absolute minimum value $f(d)$ for some $\mathrm{c}, \mathrm{d}, \in [a, b]$

Maxima and minima with differentiation

1. Find points where the gradient is zero with differentiation: i.e. solve $\frac{d}{dx} f(x) = 0$
2. What is the gradient like near these points? i.e. is it positive or negative?

3. If it changes from positive to negative ➔ ➔ local maximum

   If it changes from negative to positive ➔ ➔ local minimum



Maxima and minima with higher order differentiation

1. Find points where the gradient is zero with differentiation: i.e. solve $\frac{\mathrm{d}}{\mathrm{dx}}\mathrm{f}(\mathrm{x}) = 0$

2. What is the gradient like near these points? Is it **increasing** or **decreasing**?

   Is $\frac{d}{dx}\left(\frac{\mathrm{d}}{\mathrm{dx}}\mathrm{f}(x)\right)$ positive or negative?

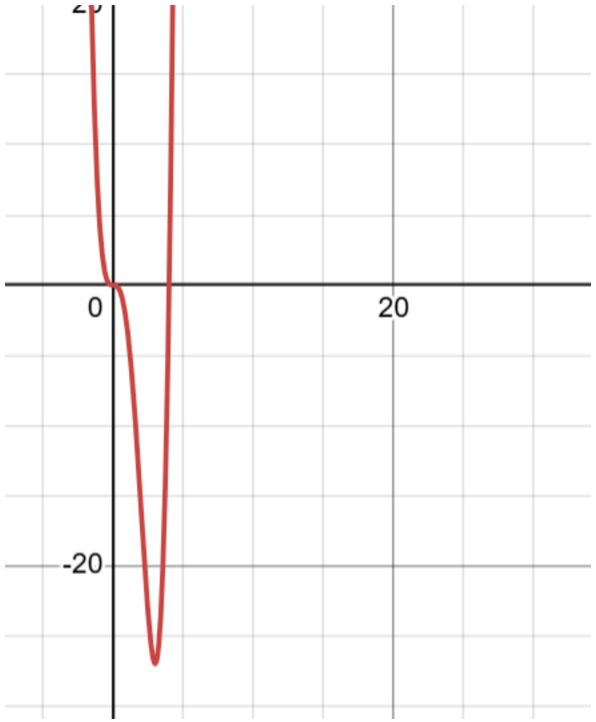
3. If $\frac{df}{dx}|_{x=c} = 0$ and $\frac{df^2}{dx^2}|_{x=c} > 0$ then f has a local minimum at c.

   If $\frac{df}{dx}|_{x=c}0$ and $\frac{df^2}{dx^2}|_{x=c} < 0$ then f has a local maximum at c.

Example: $y = x^4 - 4x^3$

- $f'(x) = 4x^3 - 12\,x^2 = 4x^2(x-3)$
- Local maxima/minima are at $x = 0, x = 3$
- $f''(x) = 12x^2 - 24x = 12x(x-2)$
- At $f''(3) = 36$. This is a local minimum.
- At $f''(0) =\ 0$. We still don't know if this is a local minimum/maximum.
- But (first derivative test) $f'(-\epsilon) = 4\epsilon^2(-\epsilon - 3)$: positive x negative= negative. So just before 0, f is decreasing…
- …$f'(\epsilon) = 4\epsilon^2(\epsilon - 3)$: positive x negative= negative. So just after 0, f is decreasing. No local minimum of maximum (or tautologically, both a minimum or a maximum at zero).

Example: $y = x^4 - 4x^3$



Multivariate differentiation

Multivariate differentiation

- Functions of two or more variables are very common.
  e.g. area of a cuboid = base x width x height.

- What happens if we change one of these variables at a time?

- To find the rate of change with one variable, we ‘zoomed in’ until the function looked like a line.
- For two variables, we’ll zoom in until the function looks like a plane. This is the **tangent plane.**

Partial derivative

Suppose we have a continuous function $f(x, y)$.

We can find its **partial derivatives** with respect to x and y:

$$f_x = \frac{\partial f}{\partial x}(x, y) = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}$$

and

$$f_y = \frac{\partial f}{\partial y}(x, y) = \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h}$$

We are keeping one variable fixed and looking at the rate of change as we vary the other variable.

This extends to functions of more variables.

Partial derivative example 1

$f(x, y) = x^2 y$

- $f_x = \lim_{h \to 0} \frac{(x+h)^2 y - x^2 y}{h} = \lim_{h \to 0} \frac{2hxy - h^2 y}{h} = 2xy$
- $f_y = \lim_{h \to 0} \frac{x^2 (y+h) - x^2 y}{h} = \lim_{h \to 0} \frac{x^2 h}{h} = x^2$

Partial derivative example 2

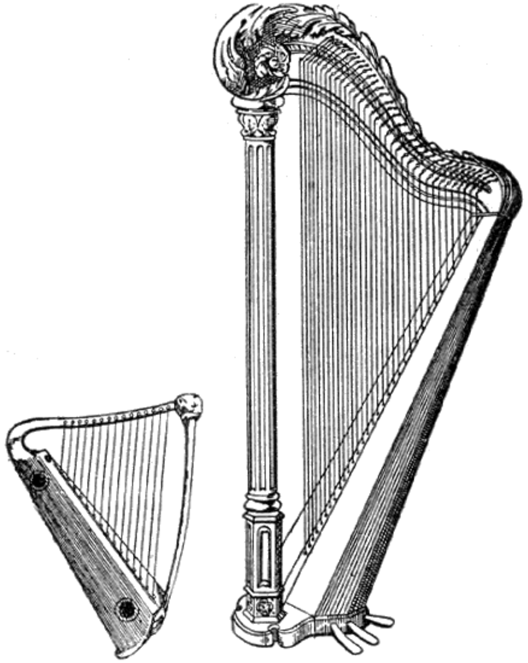
$f(x, y) = \sin\left(\frac{x}{1+y}\right)$

- $f_x = \cos\left(\frac{x}{1+y}\right)\frac{\partial}{\partial x}\left(\frac{x}{1+y}\right) = \cos\left(\frac{x}{1+y}\right)\frac{1}{1+y}$

- $f_y = \cos\left(\frac{x}{1+y}\right)\frac{\partial}{\partial y}\left(\frac{x}{1+y}\right) = -\cos\left(\frac{x}{1+y}\right)\frac{x}{(1+y)^2}$

- We used the chain rule!

Nabla: vector of partial derivatives



Harps, p. 984.

$$\nabla f(x_1, \dots, x_n) = \left(\frac{\partial f}{\partial x_1}(x_1, \dots, x_n), \dots, \frac{\partial f}{\partial x_n}(x_1, \dots, x_n)\right)$$

Example: $f(x, y) = x^2 y$

$\nabla f(x, y) = (2xy, x^2)$

Measures how f changes in the x and y directions

Derivative along a vector

Let $\mathbf{w} = (w_1, w_2)$

$$\nabla_{\boldsymbol{w}} f(x, y) = \lim_{h \to 0} \frac{f(x+hw_1, y+h\,w_2) - f(x,y)}{h}$$

- $\nabla_{\boldsymbol{w}} f(x, y) = w_1 \frac{\partial f}{\partial x}(x, y) + w_2 \frac{\partial f}{\partial y}(x, y) = \nabla f \cdot \boldsymbol{w}$
- This tells us about the rate of change in the direction $\boldsymbol{w}$. In other words, as we increase the input (x, y) by $\boldsymbol{w}$, how does our output change.
- Typically $\boldsymbol{w}$ is a unit vector
- This extends to functions of finitely many variables.
- https://math.stackexchange.com/questions/2066558/directional-derivatives-geometric-intuition?noredirect=1&lq=1

Why do we care about the directional derivative?

- If you’re walking up a hill…
- If you’re optimising in one ‘direction’
- Reverse application – to find the direction that’s increasing the fastest

Extrema: the Hessian matrix

Is my multivariate equation increasing or decreasing?

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x\, \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

(a, b) is a critical point if $f_x(a, b) = f_y(a, b) = 0$

- The eigenvalues of the Hessian at **critical points** tell us if we have a maximum or minimum.
  - Both eigenvalues negative: maximum
  - Both eigenvalues positive: minimum
  - One positive and one negative: saddlepoint
  - Zero eigenvalue: ??

Classifying Eigenvalues

- We don't actually need the eigenvalues – only their signs. We can do this by looking at certain matrix invariants:
  - Determinant: $\det H = \lambda_1 \lambda_2$
  - Trace: $\mathrm{tr}(H) = \lambda_1 + \lambda_2$

- If $\det H = 0$, then we can't classify the critical point by the Hessian
- If $\det H < 0$, then we have a positive and a negative eigenvalue → saddlepoint
- If $\det H > 0$ & $\mathrm{tr}(\mathrm{H}) > 0$, then we have two positive eigenvalues → minimum
- If $\det H > 0$ & $\mathrm{tr}(\mathrm{H}) < 0$, then we have two negative eigenvalues → maximum

Example on board

$f(x,y) = x^3 + x^2y - y^2 - 4y$

Summary

- Maxima and minima via derivatives
- Partial derivatives
- Nabla and directional derivative
- Hessian matrix
- Hessian matrix eigenvalues: tricks

Feedback

Stop:

Start:

Continue: