Some Relevant Calculus Review

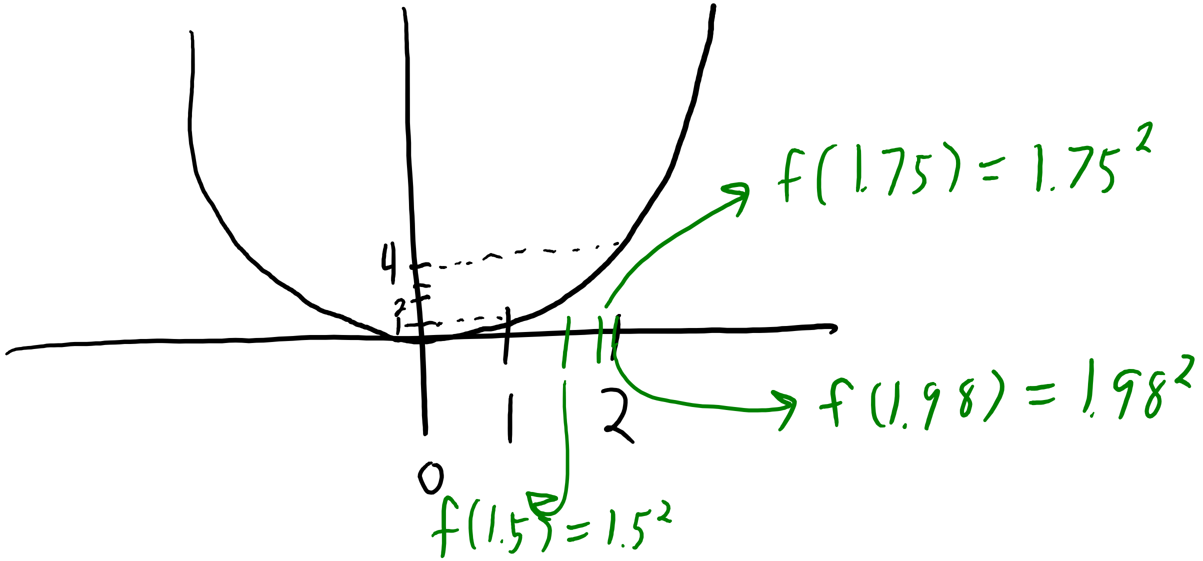
Mark Schmidt

2018

Limits (Informal Definition)

- Let ‘f’ be a function that assigns an input ‘x’ to an output f(x).
- We say ‘f’ has a limit ‘L’ at ‘c’:
  - As ‘x’ gest closer and closer to ‘c’, the value f(x) gets closer and closer to ‘L’.

- Example: the limit of $f(x)=x^2$ as ‘x’ goes to 2 is 4.



As the input gets closer to 2 the output gets closer to 4.

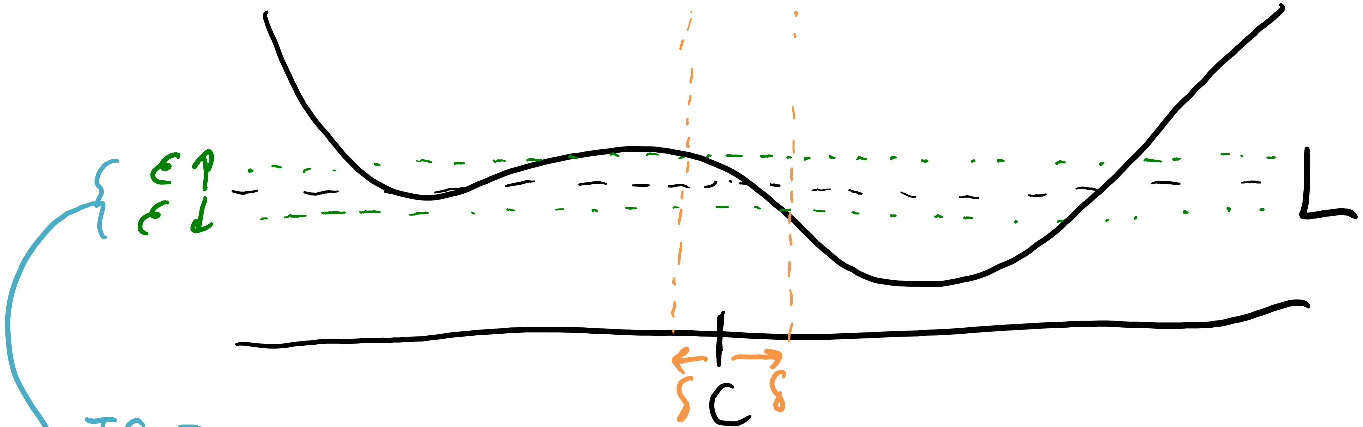
Limits (Formal Definition)

- Formally, 'f' has limit 'L' at 'c' if:
  - For all "error" values ε > 0, there exists a "distance" δ > 0 such that:
    - We have |f(x) – L| < ε for all 'x' where |x – c| < δ and x≠c.

f(x) is "close" to 'L'

'x' is "close" to 'c'

but 'x' isn't exactly 'c'



If I pick any ε then I can find a δ where all the f(x) values for x within ±δ of 'c' are within ±ε of 'L'

Limit vs. f(c)

- The standard notation for "f(x) has a limit of 'L' at 'c'" is:

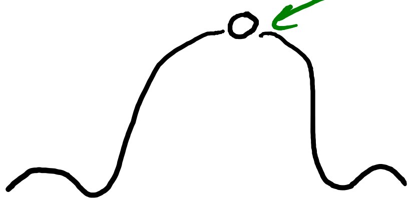
$$\lim_{x \to c} f(x) = L$$

- The limit is often simply equal to f(c), the function evaluated at 'c':

If $f(x) = x^2$ then $\lim_{x \to 2} f(x) = 4$

- However, it may not depend on f(c):

If $f(x) = \frac{\sin(x)}{x}$ the $\lim_{x \to 0} f(x) = 1$ (but $f(0)$ is not defined)



Keeps getting closer to '1' even though it doesn't exactly get there.

Continuous Functions

- We say ‘f’ is continuous at ‘c’ if f(c) equals the limit of ‘f’ at ‘c’:

$$\lim_{x \to c} f(x) = f(c)$$

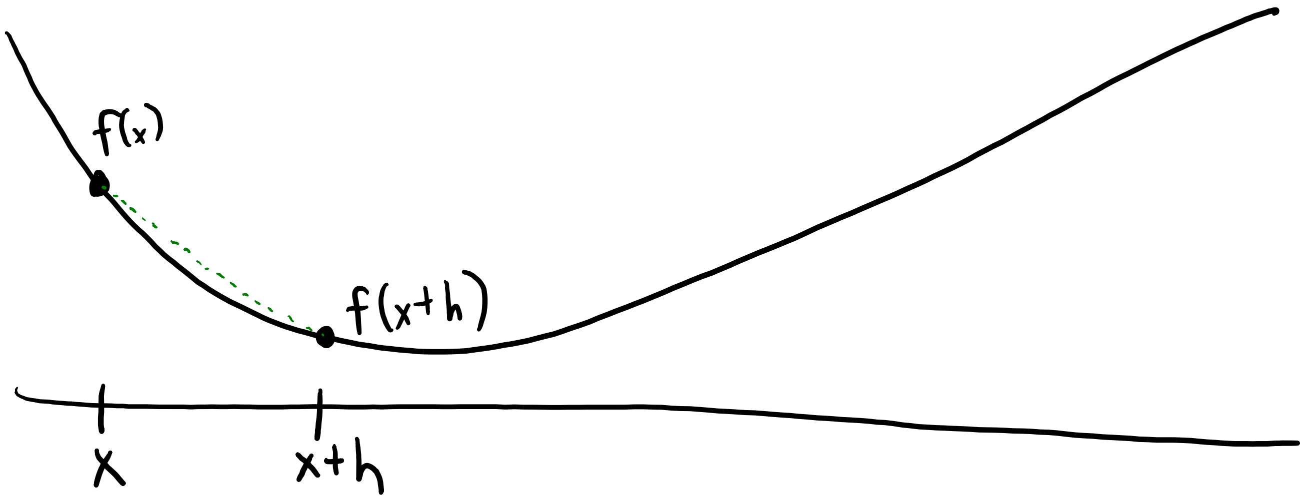
  – We say it’s discontinuous at ‘c’ if this equality is false (doesn’t equal limit).

- We’ll say a “function is continuous” if it’s continuous for all inputs.
  – This roughly means that “small changes in ‘x’ lead to small changes in f(x)”.

- Most simple functions like polynomials are continuous.
  - The composition of continuous functions will also be continuous.

Average Rate of Change

- Consider the interval from ‘x’ to ‘x+h’ for some function ‘f’ and h > 0:



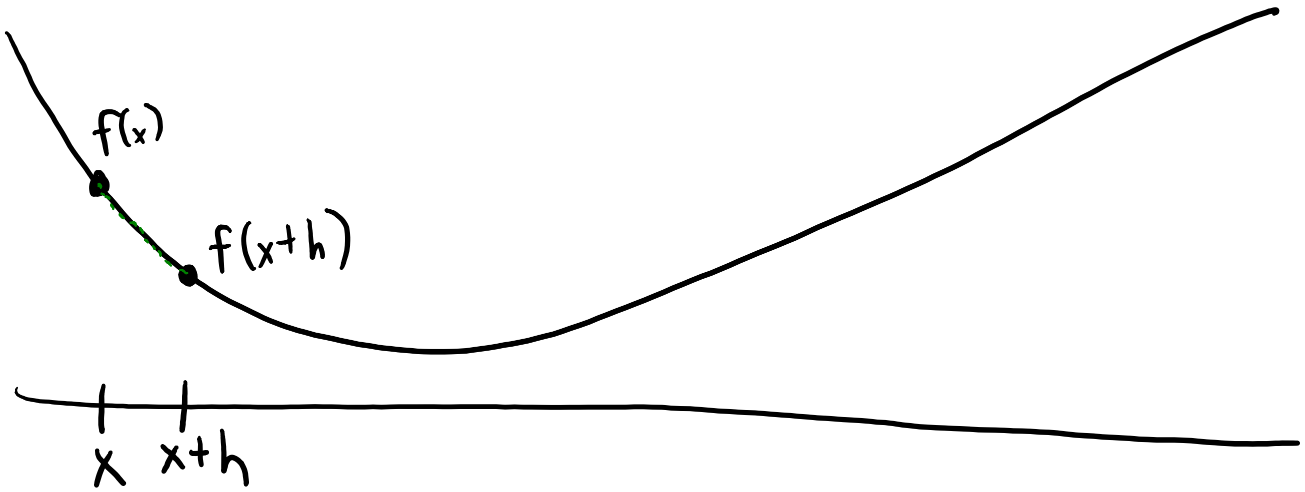
- The “average rate of change” of the function over the interval is:

$$\frac{f(x+h) - f(x)}{h}$$

- For linear functions, f(x) = ax + b, this gives the slope ‘a’ for any x and h.

Derivative

- Get more accurate measure of instantaneous change with smaller 'h':



- The derivative f'(x) is the limit of the rate of change as 'h' goes to zero:

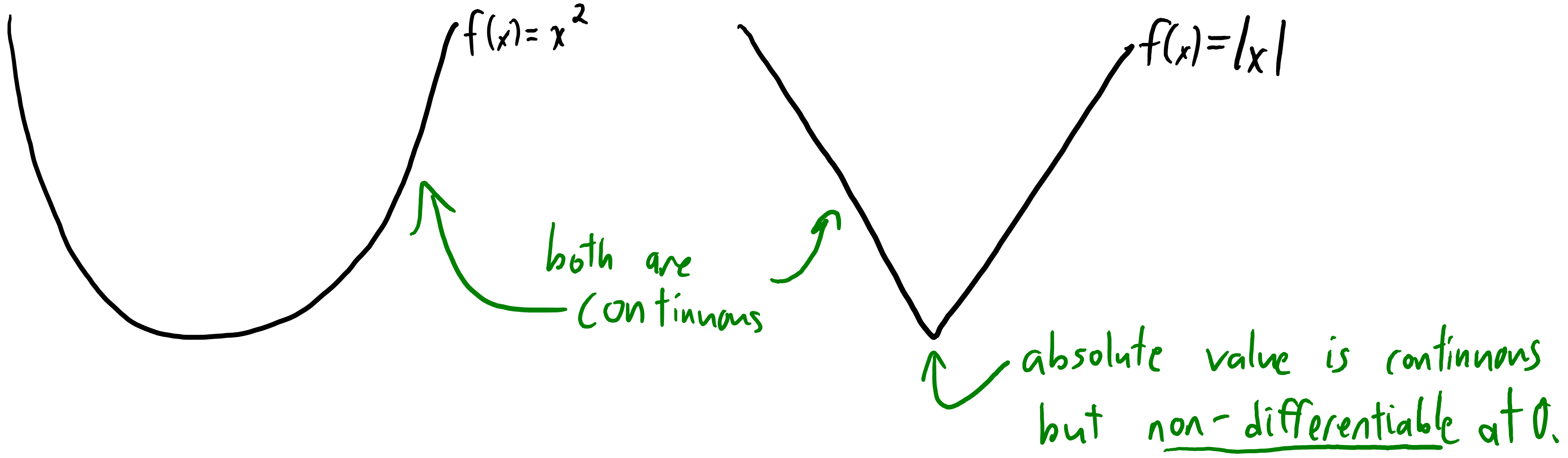
$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

- For linear functions, f(x) = ax + b, the derivative is the slope, f'(x) = a.

Derivatives and Continuity

- We say that ‘f’ is differentiable at ‘c’ if the derivative exists at ‘c’.

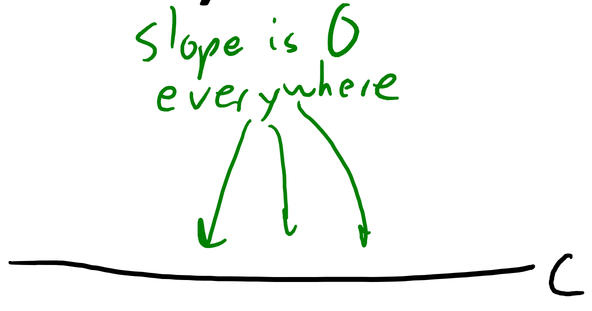
- If ‘f’ is differentiable at ‘c’, it must be continuous.
  - But ‘f’ can be continuous without being differentiable.



Common Derivatives (Polynomials)

- Derivative of constant function is 0:

$$\text{If } f(x) = c \text{ then } f'(x) = 0$$



- Power rule for derivative of simple polynomial:

$$\text{If } f(x) = x^p \text{ then } f'(x) = p x^{p-1}$$

  – Example: if $f(x) = x^2$ then $f'(x) = 2x$.

- Multiplying f(x) by a constant changes derivative by a constant:
  – Example: if $f(x) = 2x^2$ then $f'(x) = 4x$.

Common Derivatives (Exponential and Logarithm)

- Derivatives can be computed term-wise:

$$\text{If } f(x) = g(x) + h(x) \text{ then } f'(x) = g'(x) + h'(x)$$

- The derivative of the exponential function (e^x) is itself:

$$\text{If } f(x) = \exp(x) \text{ then } f'(x) = \exp(x)$$

- The derivative of the logarithm function (base 'e') is the reciprocal:

$$\text{If } f(x) = \log(x) \text{ then } f'(x) = \frac{1}{x} \quad (\text{for } x > 0)$$

- Note that we're defining exp(x) and log(x) so that log(exp(x))=x.

Common Derivatives (Composition)

- The chain rule lets us take derivatives of compositions:

$$\text{If } f(x) = g(h(x)) \quad \text{then } f'(x) = g'(h(x))h'(x)$$

  – Example: if $f(x) = \exp(x^2)$, then $f'(x) = \exp(x^2)2x$.

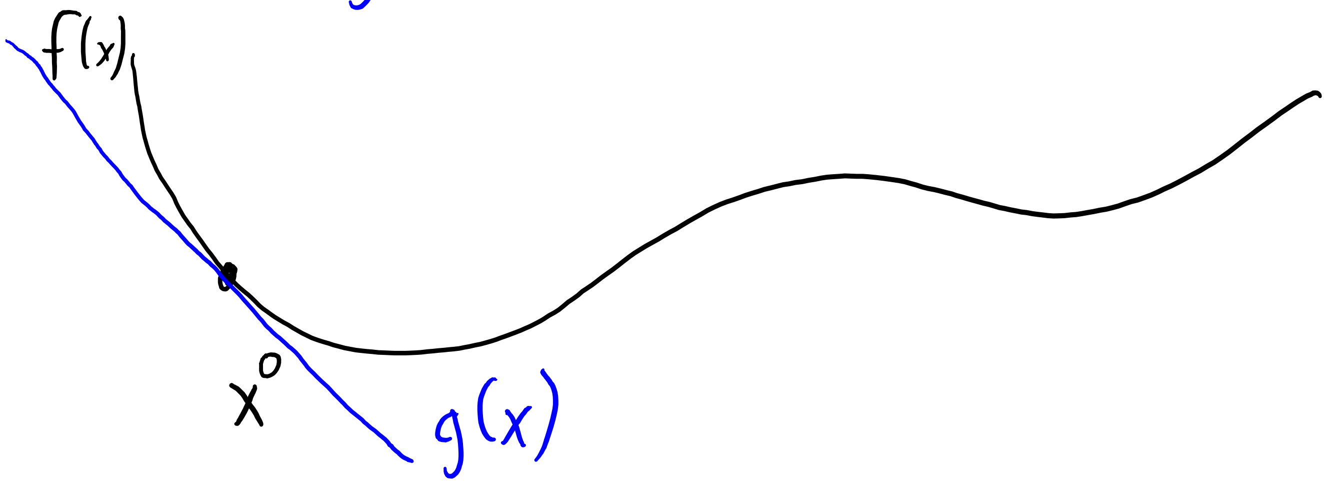
- A 3Blue1Brown video with intuition for common derivatives.

Tangent Line

- If 'f' is differentiable at a point x^0, the tangent line is given by:

$$g(x) = f(x^0) + f'(x^0)(x - x^0)$$

the line with slope $f'(x^0)$ and y-intercept $f(x^0) - f'(x^0)x^0$



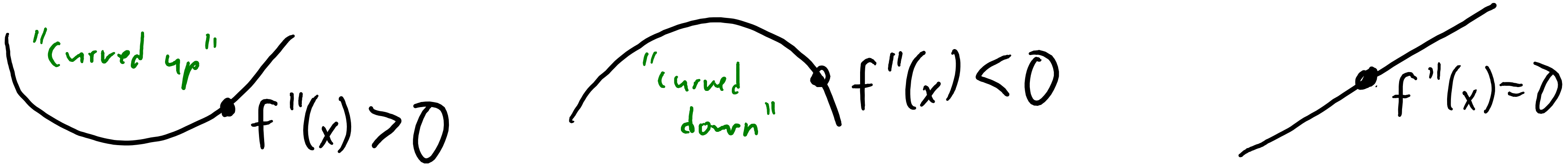
- The tangent line 'g' is the unique line such that at x^0 we have:
  - Same function value: $g(x^0) = f(x^0)$.
  - Same derivative value: $g'(x^0) = f'(x^0)$.
- We often use the tangent line as a "local" approximation of 'f'.

Higher-Order Derivatives

- Second derivative $f''(x)$ is the derivative of the derivative function.
  - Gives "instantaneous rate of change" of the derivative.

  - Example: if $f(x) = x^3$ then $f'(x) = 3x^2$ and $f''(x) = 6x$.

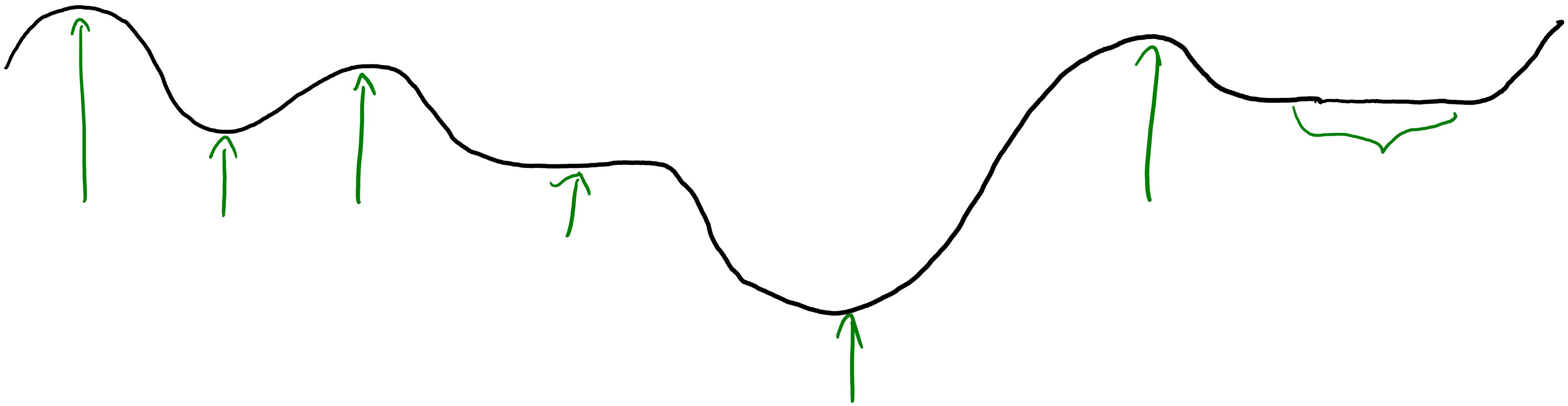
- Sign of second derivative: whether function is "curved" up or down.



- We if we take the derivative 'k'-times and the derivatives exist, we say that 'f' is k-times differentiable.

Stationary/Critical Points

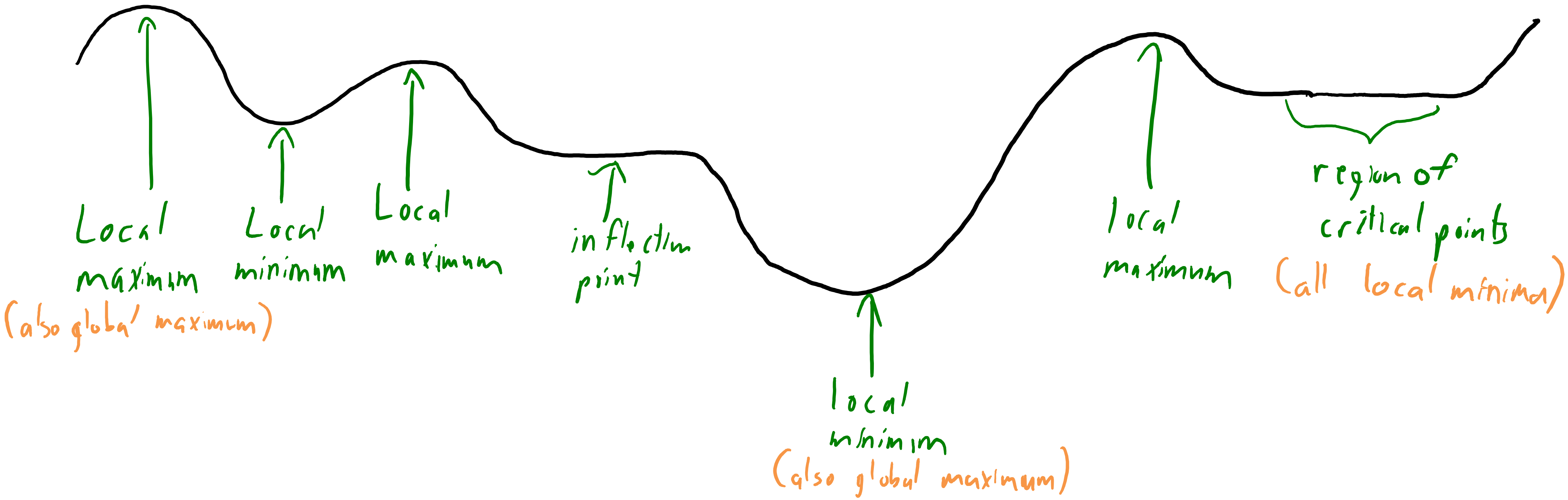
- An ‘x’ with f’(x)=0 is called a stationary point or critical point.
  - The slope is zero so the tangent line is “flat”.



Critical points

Derivative Test and Local Minima

- An ‘x’ with f’(x)=0 is called a stationary point or critical point.
  - The slope is zero so the tangent line is “flat”.



  - If f’(x) = 0 and f’’(x) >= 0, we say that ‘x’ is a local minimum.
    - “Close to ‘x’, there is no larger value of f(x)”.

Summation Notation and Infinite Summations

- For a sequence of variables x_i, recall summation notation:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

- We can write sum of infinite sequence of variables as a limit:

$$\lim_{n \to \infty} \sum_{i=1}^{n} x_i$$

- Another common way to write infinite summations is:

$$\sum_{i=1}^{\infty} x_i$$

Bounding Summations

- Some useful facts about summations:

$$\sum_{i=1}^{t} \frac{1}{i^2} = O(1) \qquad \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty \qquad \text{"converges"}$$

$$\sum_{i=1}^{t} \frac{1}{i} = O(\log t) \qquad \sum_{i=1}^{\infty} \frac{1}{i} = \infty \qquad \text{"diverges"}$$

$$\sum_{i=1}^{t} \frac{1}{\sqrt{i}} = O(\sqrt{t}) \qquad \sum_{i=1}^{\infty} \frac{1}{\sqrt{i}} = \infty \qquad \text{"diverges faster"}$$

Partial Derivatives

- Multivariate functions have more than one variable:

$$f(x_1, x_2, x_3) = \underbrace{a_1 x_1 + a_2 x_2 + a_3 x_3 + b}_{\text{"multivariate linear"}}$$

- Partial derivative: derivative of one variable, with all others fixed:

$$\underbrace{\frac{\partial}{\partial x_1}}_{\text{"partial derivative with respect to } x_1\text{"}}\left[f(x_1, x_2, x_3)\right] = a_1$$

$$\frac{\partial}{\partial x_2}\left[f(x_1, x_2, x_3)\right] = a_2$$

$$\frac{\partial}{\partial x_3}\left[f(x_1, x_2, x_3)\right] = a_3$$

Gradient

- Gradient is a vector containing partial derivative ‘i’ in position ‘i’.

$$\begin{array}{cl} \nabla f(x_1, x_2, x_3, x_4) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \\ \frac{\partial f}{\partial x_4} \end{bmatrix} & \begin{array}{l} \leftarrow \text{position } 1 \\ \leftarrow \text{position } 2 \\ \leftarrow \text{position } 3 \\ \leftarrow \text{position } 4 \end{array} \end{array}$$

- Example:

Function

$$f(x_1, x_2, x_3) = a_1 x_1 + a_2 x_2 + a_3 x_3 + b$$

Partial derivatives

$$\frac{\partial}{\partial x_1}\left[f(x_1, x_2, x_3)\right] = a_1$$

$$\frac{\partial}{\partial x_2}\left[f(x_1, x_2, x_3)\right] = a_2$$

$$\frac{\partial}{\partial x_3}\left[f(x_1, x_2, x_3)\right] = a_3$$

Gradient

$$\nabla f(x_1, x_2, x_3) = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Multivariate Quadratic

- Multivariate quadratic is a multi-variable degree-2 polynomial:

$$f(x_1, x_2) = \frac{1}{2} a_{11} x_1^2 + a_{12} x_1 x_2 + \frac{1}{2} a_{22} x_2^2 + b_1 x_1 + b_2 x_2 + c$$

$$\frac{\partial f}{\partial x_1} = a_{11} x_1 + a_{12} x_2 + b_1$$

$$\frac{\partial f}{\partial x_2} = a_{12} x_1 + a_{22} x_2 + b_2$$

$$\nabla f(x_1, x_2) = \begin{bmatrix} a_{11} x_1 + a_{12} x_2 + b_1 \\ a_{12} x_1 + a_{22} x_2 + b_2 \end{bmatrix}$$

Matrix Notation for Linear Functions (Objective)

- Common ways to write a multivariate linear function:

$$f(x_1, x_2, x_3) = a_1x_1 + a_2x_2 + a_3x_3 + b$$

$$= \sum_{i=1}^{3} a_i x_i + b \qquad \text{(summation notation)}$$

$$= a^T x + b \qquad \text{(matrix notation)}$$

- The last line uses matrix notation, defining vectors 'a' and 'x' as:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad \text{so that} \quad a^T x = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \sum_{i=1}^{3} a_i x_i$$

Matrix Notation for Linear Functions (Gradient)

- So we can write a multivariate linear function as:

$$f(x) = \underbrace{a^T x}_{\sum_{j=1}^{3} a_j x_j} + b$$

- We can also write the gradient in matrix notation:

$$\nabla f(x) = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = a$$

Matrix Notation for Quadratic Functions (Objective)

- Common ways to write a multivariate quadratic function:

$$f(x_1, x_2) = \frac{1}{2} a_{11} x_1^2 + a_{12} x_1 x_2 + \frac{1}{2} a_{22} x_2^2 + b_1 x_1 + b_2 x_2 + c$$

$$= \frac{1}{2} \sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} x_i x_j + \sum_{i=1}^{2} b_i x_i + c \quad \text{(summation notation)}$$

Assuming $a_{12} = a_{21}$ (symmetric)

$$= \frac{1}{2} x^T A x + b^T x + c \quad \text{(matrix notation)}$$

- Using vectors 'b' and 'x' and matrix 'A', and matrix multiplication:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{so } b^T x = \sum_{i=1}^{2} b_i x_i \text{ and } x^T A x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \sum_{j=1}^{2} a_{1j} x_j \\ \sum_{j=1}^{2} a_{2j} x_j \end{bmatrix} = \sum_{i=1}^{2} x_i \sum_{j=1}^{2} a_{ij} x_j$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} x_i x_j$$

Matrix Notation for Quadratic Functions (Gradient)

- So we can write a multivariate quadratic function as:

$$f(x) = \frac{1}{2}\underbrace{x^T A x}_{\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij} x_i x_j} + \underbrace{b^T x}_{\sum_{i=1}^{d} b_i x_i} + c$$

- We can also write the gradient in matrix notation (symmetric 'A'):

$$\nabla f(x) = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + b_1 \\ a_{12}x_1 + a_{22}x_2 + b_2 \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}}_{A}\underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}_{b} = Ax + b$$

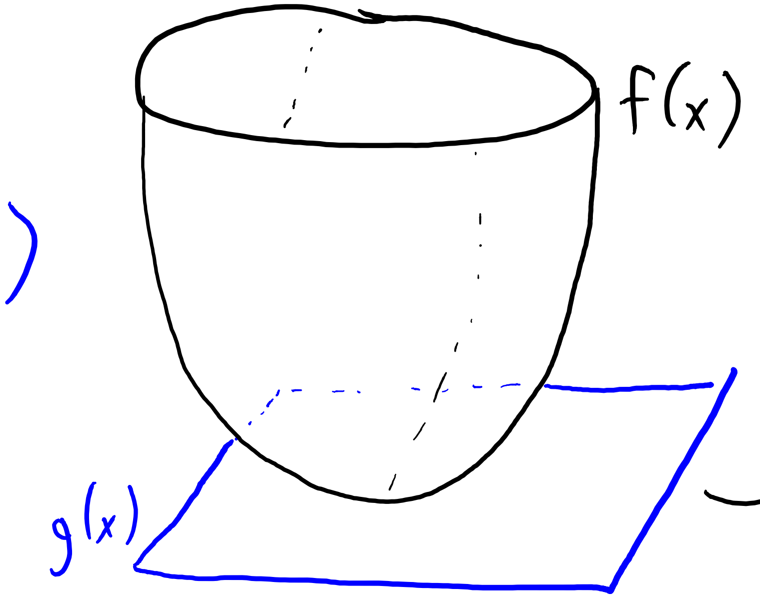
- In non-symmetric case:

$$\nabla f(x) = \frac{1}{2}(A + A^T)x + b$$

Tangent [Hyper-]Plane

- If 'f' is differentiable at a point x^0, the tangent hyper-plane is given by:

$$g(x) = f(x^0) + \nabla f(x^0)^T (x - x_0)$$



Here $\nabla f(x) = 0$ so it's called a stationary point.

- The tangent hyper-plane 'g' is the unique hyper-plane such that at x^0:
  - Same function value: $g(x^0) = f(x^0)$.
  - Same partial derivative values: $\frac{\partial}{\partial x_i} g(x^0) = \frac{\partial}{\partial x_i} f(x^0)$
- We often use the tangent hyper-plane as a "local" approximation of 'f'.