Last time:

1st order Taylor series approx (of f around the point x.

$$\boxed{\begin{array}{l}\text{If } f: \mathbb{R}^n \to \mathbb{R} \text{ is twice continuously differentiable} \\ f(y) = f(x) + \nabla f(x)^T (y-x) \\ \quad + \frac{1}{2}(y-x)^T \nabla^2 f\big|_{\xi} (y-x)\end{array}}$$

here ξ is a point on the line segment connecting x & y

so $\xi = ty + (1-t)x$

where $t \in (0,1)$

For 1 variable

$$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2} f''(\xi)(y-x)^2$$

Some Equivalent conditions for convexity

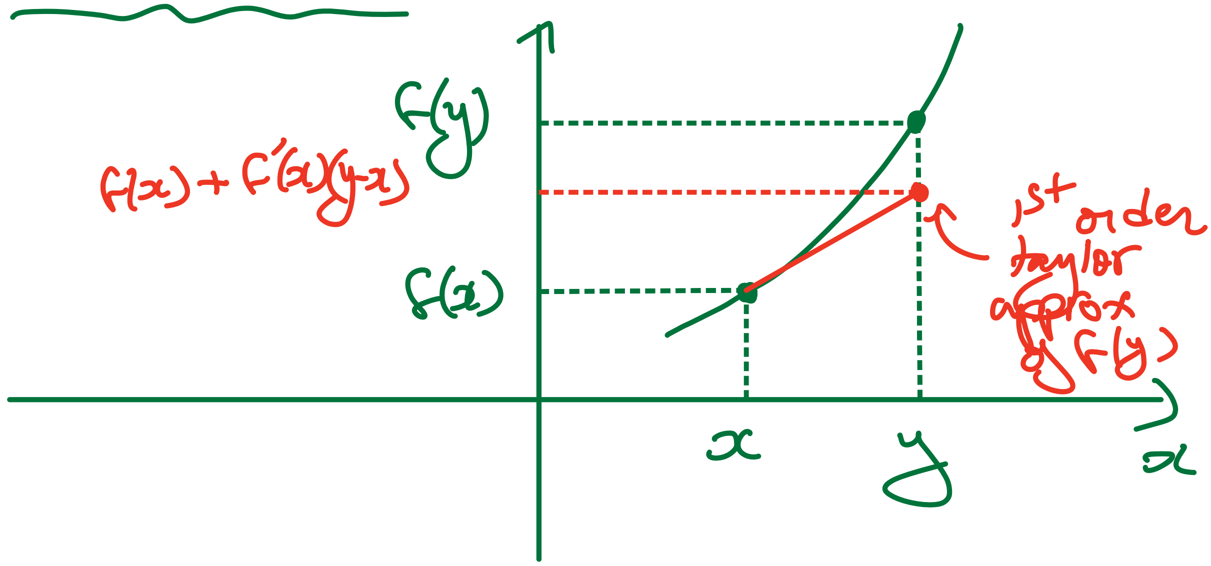
Theorem: If $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable then f is convex if and only if $\forall x \,\&\, y \in \mathbb{R}^n$ we have

$$\underbrace{f(y) \geq f(x) + \nabla f(x)^T (y-x)}$$

Important!

$f(x) + \nabla f(x)^T (y-x)$: 1st order taylor approx. of $f(y)$.

Illustration: $f : \mathbb{R} \to \mathbb{R}$



general version of single var. calc. result

Theorem: If f is twice continuously differentiable then f is convex if and only if $\forall x$:

$\nabla^2 f(x)$ is positive semidefinite.

Important!

Recall: • $A \in \mathbb{R}^{n \times n}$ is positive semidefinite (PSD) if and only if

$$x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n$$

• $\boxed{A \text{ is PSD}} \iff \boxed{\text{eigenvalues of } A \geq 0}$

Recall : λ is an eigenvalue of A if

$Av \quad \lambda v$ for some vector v (called an eigenvector)

Example : Let $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

Consider $x \in \mathbb{R}^3$, $x = (x_1, x_2, x_3)$

We'll check $x^T A x = [x_1 \; x_2 \; x_3] \, A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

$$= [x_1 \;\; x_2 \;\; x_3] \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= [2x_1 - x_2 \quad -x_1 + 2x_2 - x_3 \quad -x_2 + 2x_3] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= 2x_1^2 - x_1x_2 - x_1x_2 + 2x_2^2 - x_2x_3 - x_2x_3 + 2x_3^2$$

$$= x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2 \geq 0$$

$\forall x \Rightarrow A$ is PSD.

Exercise: Repeat the problem but this time by calculating the eigenvalues of A.

Example: Check whether $f: \mathbb{R}^3 \to \mathbb{R}$ given by $f(a,b,c) = a^4 + b^2 + c^2$ is convex.

Solution: as f is twice continuously diff., it is enough to check whether $\nabla^2 f$ is PSD

$$\nabla f(a,b,c) = (4a^3, 2b, 2c)$$

$$\Rightarrow \nabla^2 f(a,b,c) = \begin{pmatrix} 12a^2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

which is a PSD matrix no matter what a is (i.e. $\forall a$) so f is convex.

Some more on convexity & derivatives:

Defn: A mapping $g : \mathbb{R}^n \to \mathbb{R}^n$ is called <u>monotone</u> if $\forall x, y \in \text{domain}(g)$

$$\langle g(x) - g(y), x - y \rangle \geq 0.$$

Example: $g : \mathbb{R}^n \to \mathbb{R}^n$

$$g(x) = 2x$$

Then $\forall x, y \in \mathbb{R}^n$

$$\langle g(x) - g(y), x - y \rangle =$$

$$\langle 2(x) - 2(y), x - y \rangle =$$

$$2 \|x - y\|^2 \geq 0$$

$\Rightarrow g$ is monotone

Theorem: A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if ∇f is monotone

Proof: (I) convex $\Longrightarrow$ gradient monotone:

By a previous theorem (on page 2)

f is convex $\Longrightarrow$

$$\begin{cases} f(y) \geq f(x) + \nabla f(x)^T (y-x) \\ f(x) \geq f(y) + \nabla f(y)^T (x-y) \end{cases}$$

adding the 2 inequalities

$$\cancel{f(x)} + \cancel{f(y)} \geq \cancel{f(x)} + \cancel{f(y)} + \nabla f(x)^T (y-x) + \nabla f(y)^T (x-y)$$

$$\Longrightarrow 0 \geq \left[\nabla f(x) - \nabla f(y)\right]^T (y-x)$$

$$\Longrightarrow \left[\nabla f(x) - \nabla f(y)\right]^T (x-y) \geq 0$$

$\hookrightarrow$ multiply both sides by -1 and reverse the inequality

$\Longrightarrow \nabla f$ is monot…

$\Rightarrow \nabla f$ is monotone

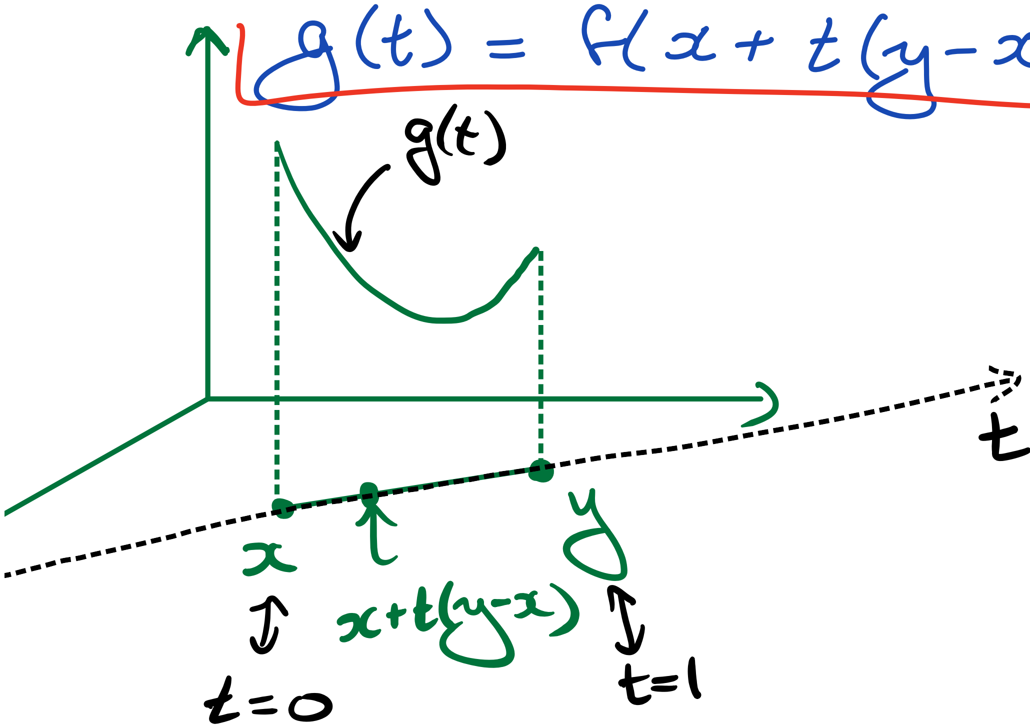
(II) gradient monotone $\Rightarrow$ convex

The idea of this proof is to show that monotonicity of ∇f implies

$$f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

To do this we'll parametrize the line segment between x & y, and define a new function $g: \mathbb{R} \to \mathbb{R}$

$$\boxed{g(t) = f(x + t(y-x))}$$



$$g'(t) = \nabla f(x + t(y-x))^T (y-x)$$

↑ Chain rule from multi-variable calc.

Now, let's use the fact that the gradient is monotone, i.e.

$$\langle \nabla f(z) - \nabla f(x),\ z - x \rangle \geq 0$$
$\forall z, x$

Picking $z = x + t(y-x)$ we get

$$\langle \nabla f(\overbrace{x + t(y-x)}^{z}) - \nabla f(x),\ \overbrace{x + t(y-x)}^{z} - x \rangle \geq 0$$

$$\Leftrightarrow \langle \nabla f(x + t(y-x)) - \nabla f(x),\ t(y-x) \rangle \geq 0$$

$$\Leftrightarrow \quad t[g'(t) - g'(0)] \geq 0$$

$$\Rightarrow \quad g'(t) \geq g'(0)$$

$$\Rightarrow \int_0^1 g'(t)\,dt \geq \int_0^1 g'(0)\,dt$$

$$\Rightarrow g(1) - g(0) \geq g'(0)$$

$$\begin{array}{ccccc} g(1) & \geq & g(0) & + & g'(0) \\ \| & & \| & & \| \\ f(y) & & f(x) & & \nabla f(x)^T(y-x) \end{array}$$

$$\Rightarrow f(y) \geq f(x) + \nabla f(x)^T(y-x)$$

i.e. f is convex !

$$g'(t) = \nabla f(x + t(y-x))^T (y-x)$$

$$\Rightarrow g'(0) = \underline{\nabla f(x)^T (y-x)}$$

$$t\left(g'(t) - g'(0)\right) =$$

$$\left[\nabla f(x + t(y-x)) - \nabla f(x)\right]^T t(y-x)$$

$$= \langle \nabla f(x + t(y-x)) - \nabla f(x), t(y-x) \rangle$$