<u>Theorem</u> $\mathcal{S}$: convex set, $f: \mathcal{S} \to \mathbb{R}$ convex function

(i) If $x^*$ is local min., then $x^*$ is global min.

(ii) If $f$ is strictly convex, then min. is unique

(iii) Replace min by max in (i), (ii) for concave $f$

<u>Proof</u> (i) Suppose $x^* \in \mathcal{S}$ is local min. but not global

Then, $\exists\, y \in \mathcal{S}$ s.t. $f(y) < f(x^*)$

$$f(\underbrace{\alpha y + (1-\alpha) x^*}_{\in \mathcal{S}}) \leq \alpha f(y) + (1-\alpha) f(x^*), \quad \forall \alpha \in (0,1)$$

$$< \alpha f(x^*) + (1-\alpha) f(x^*) = f(x^*) \quad \forall \alpha \in (0,1)$$

Thus

$$f(x^* + \alpha(y - x^*)) < f(x^*) \quad \forall \alpha\ (0,1)$$

Since this is true for $\alpha > 0$ however small

$x^*$ cannot be a local min. $\Rightarrow\Leftarrow$

(ii) Suppose $y$ is another min. Then $f(y) = f(x^*)$

$$f\left(\tfrac{1}{2} x^* + \tfrac{1}{2} y\right) \underset{\text{strict convexity}}{<} \tfrac{1}{2} f(y) + \tfrac{1}{2} f(x^*) = f(x^*)$$

Since $\frac{1}{2} x^* + \frac{1}{2} y \in \mathcal{S}$, $x^*$ is not a min. $\Rightarrow\Leftarrow$

(iii) Similar to (i), (ii) with $\leq$, $<$ replaced by $\geq$, $>$.

# First Derivative Characterization of Convexity

**Theorem** (i) Let $\mathcal{S}$ be a convex set, and $f: \mathcal{S} \to \mathbb{R}$ be a continuously differentiable function

$$f \text{ is convex on } \mathcal{S} \iff f(y) \geq f(x) + \nabla f(x)^T (y-x), \quad \forall x, y \in \mathcal{S}$$

(ii) Inequality is strict for strict convexity

**Proof** (i) "$\Rightarrow$" part

$$f(x + \alpha(y-x)) = f((1-\alpha)x + \alpha y)$$

$$\xrightarrow{\text{convexity}} \leq (1-\alpha) f(x) + \alpha f(y), \quad \alpha \in (0,1)$$

$$\Rightarrow \quad \frac{f(x+\alpha(y-x)) - f(x)}{\alpha} \leq f(y) - f(x)$$

limit as $\alpha \to 0 \quad \Rightarrow \quad \nabla f(x)^T (y-x) \leq f(y) - f(x)$

"$\Leftarrow$" part. For $\alpha \in [0,1]$, let $z = \alpha x + (1-\alpha) y$

$$\alpha \cdots \quad f(z) + \nabla f(z)^T (x-z) \leq f(x)$$

$$(1-\alpha) \cdots \quad f(z) + \nabla f(z)^T (y-z) \leq f(y)$$

$$f(z) + \nabla f(z)^T \underbrace{(\alpha(x-z) + (1-\alpha)(y-z))}_{=0} \leq \alpha f(x) + (1-\alpha) f(y)$$

Thus $f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$

(ii) Replace "$\leq$" by "$<$" in (i)

## Taylor's Theorem (review)

For $f : \mathbb{R} \to \mathbb{R}$,

$i^{th}$ derivative of $f$

$$f(y) = \sum_{i=0}^{n} \frac{f^{(i)}(x)}{i!}(y-x)^i + \frac{f^{(n+1)}(z)}{(n+1)!}(y-x)^{n+1}$$

for some $z$ between $x$ and $y$, equivalently

$$z = \alpha x + (1-\alpha) y \quad \text{for some } \alpha \in [0,1]$$

Special case of $n=0$ gives Mean Value Theorem:

$$f(y) = f(x) + f'(z)(y-x).$$

### Approximation

$$f(y) = \sum_{i=1}^{n} \frac{f^{(i)}(x)}{i!}(y-x)^i + O((y-x)^n)$$

## Multivariate Generalization $\quad f : \mathbb{R}^n \to \mathbb{R}$

First-order: $f(y) = f(x) + \nabla f(z)^T (y-x)$

for $z = \alpha x + (1-\alpha) y$ for some $\alpha \in [0,1]$

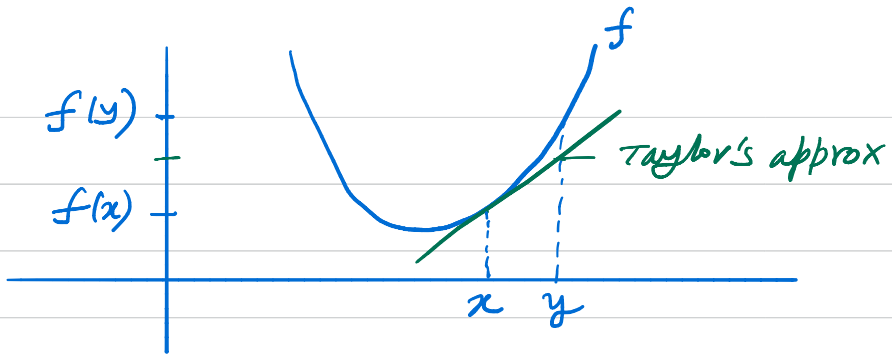
Approx.: $f(y) = f(x) + \nabla f(y)^T (y-x) + O(\|y-x\|)$

---

First-order characterization (previous page):

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

Taylor's approx. underestimates a convex function!




## Second order Taylor Expansion (Multivariate)

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2} (y-x)^T \nabla^2 f(z) (y-x) \quad -(*)$$

for $z = \alpha x + (1-\alpha) y$ for some $\alpha \in [0,1]$.

approx. $= f(x) + \nabla f(x)^T (y-x) + \frac{1}{2} (y-x)^T \nabla^2 f(x) (y-x) + o(\|y-x\|^2)$

---

**Theorem** Let $S$ be convex set, and $f: S \to \mathbb{R}$ twice cont. diff.

(i) $\nabla^2 f(x) \succeq 0$, $\forall x \in S \Rightarrow f$ is convex on $S$

(ii) $\nabla^2 f(x) \succ 0$, $\forall x \in S \Rightarrow f$ is strictly convex on $S$

(iii) $\succeq$ becomes $\preceq$ and $\succ$ becomes $\prec$ for concave $f$.

**Proof** (i) Suppose that $\nabla^2 f(x) \succeq 0 \quad \forall x \in S$

For $y \in S$, $\alpha \in [0,1]$, $z = \alpha x + (1-\alpha) y \in S \Rightarrow \nabla^2 f(z) \succeq 0$

Plugging into $(*)$ we get,

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

$\Rightarrow f$ is convex by first-order condition

(ii) and (iii) follow similarly.

**Theorem** If $f$ is twice continuously differentiable and convex over open set $S$, then $\nabla^2 f(x) \succeq 0, \forall x \in S$.

**Proof** (by contradiction). Suppose $\nabla^2 f(\cdot)$ is not PSD for all points in $S$. Then $\exists u \in \mathbb{R}^n, y \in S$ s.t.

$$u^T \nabla^2 f(y) u < 0$$

By continuity of $\nabla^2 f$, and $S$ being open, if $\|u\|$ is small enough, $y - \alpha u \in S$ and

$$u^T \nabla^2 f(y - \alpha u) u < 0 \quad \forall \alpha \in [0,1]$$

Now let $x = y - u$. Then we have

$$(y-x)^T \nabla^2 f(\underbrace{y - \alpha(y-x)}_{\alpha x + (1-\alpha) y \in S})(y-x) < 0, \quad \forall \alpha \in [0,1]$$

Plugging into (*), we get

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

$$\Rightarrow f \text{ is not convex} \Rightarrow\Leftarrow.$$

■

---

Note: $f$ strictly convex on $S$ $\not\Rightarrow$ $\nabla^2 f(x) \succ 0 \quad \forall x \in S$

Counter-example: $f(x) = x^4$ (strictly convex)

$$\frac{d^2}{dx^2} f(x) = 12x^2 \;(= 0 \text{ at } x = 0)$$

Example 1) Affine function: $f(x) = a^T x + b$, $x \in \mathbb{R}^n$

$\nabla f(x) = a \qquad \nabla^2 f(x) = 0 \quad \forall x$

$\Rightarrow$ $f$ is both convex and concave, but not strict

Note that we cannot conclude that $f$ is not strictly convex/concave from previous theorems.
But if we assume $f$ is strictly convex, then
$b = f(0) = f(\frac{1}{2}x - \frac{1}{2}x) < \frac{1}{2}f(x) + \frac{1}{2}f(-x) = b \Rightarrow\Leftarrow$

Similarly if we assume $f$ is strictly concave, we get $b > b$

2) Quadratic function: $f(x) = \frac{1}{2}x^T Q x + b^T x + c$, $x \in \mathbb{R}^n$
$Q$ symmetric (w.l.o.g.)

$\nabla f(x) = Qx + b, \qquad \nabla^2 f(x) = Q.$

1) $Q \succeq 0 \Leftrightarrow f$ is convex
2) $Q \succ 0 \Leftrightarrow f$ is strictly convex
3) $Q \preceq 0 \Leftrightarrow f$ is concave
4) $Q \prec 0 \Leftrightarrow f$ is strictly concave

1) and 3), and "$\Rightarrow$" parts of 2) and 4) follow from previous results. To show "$\Leftarrow$" part of 2) and 4), suppose $f$ is strictly convex, and $Q$ is $\succeq 0$ but not $\succ 0$ then $Q$ must have a $0$ eigenvalue $\Rightarrow$ There exists $x$ s.t. $Qx = 0$. Then

$c = f(0) = f(\frac{1}{2}x - \frac{1}{2}x) = \frac{1}{2}f(x) + \frac{1}{2}f(-x)$

$\Rightarrow$ $f$ not strictly convex $\Rightarrow\Leftarrow$

## Special Case

$$f(x_1,x_2,x_3) = 2x_1^2 + x_1x_2 + 2x_2x_3 + \frac{1}{2}x_3^2 + x_1 + 2x_2 + 3$$

$$f(x) = [x_1 \ x_2 \ x_3] \begin{bmatrix} 2 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & 1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + [1 \ 2 \ 0] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + 3$$

$$= \frac{1}{2} x^T Q x + b^T x + c$$

with $Q = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ $c = 3$

Recall that matrix A is PSD iff all principal minors $\geq 0$

$$\det\left(\begin{bmatrix} 0 & 2 \\ 2 & 1 \end{bmatrix}\right) = -4 < 0 \Rightarrow Q \text{ is not PSD}$$

Also, $-Q$ has principal minor $\det([-4]) < 0$

$\Rightarrow Q$ is not NSD

$Q$ is neither PSD nor NSD $\Rightarrow f$ is neither convex nor concave

---

**Result** If $f$ is a convex function over convex set $S$

$$\nabla f(x^*) = 0 \Rightarrow x^* \text{ is a global min.}$$

**Proof** $f$ convex $\Rightarrow f(y) \geq f(x^*) + \cancel{\nabla f(x^*)^T}^{=0}(y - x^*) \quad \forall y \in S$

$$\Rightarrow f(y) \geq f(x^*) \quad \forall y \in S$$

$$\Rightarrow x^* \text{ is global min.}$$

## Finding the Optimum

If $f$ is convex, and continuously differentiable then we can find a global min. by solving

$$\nabla f(x) = 0$$

If $f$ is strictly convex, there is only one solution $x^*$

(for concave function, replace min. by max.)

**Example** $f(x_1, x_2) = x_1^2 + x_1 x_2 + 2x_2^2 - x_1$

$$= x^T \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix} x + [-1 \ 0] x$$

$$\nabla f(x) = 2 \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix} x + \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

$$\nabla^2 f(x) = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \succ 0$$

$$\nabla f(x) = 0 \quad \Rightarrow \quad \begin{aligned} 2x_1 + x_2 - 1 &= 0 \\ x_1 + 4x_2 &= 0 \end{aligned}$$

$$\Rightarrow \quad x_2^* = -\frac{1}{7}, \ x_1^* = \frac{4}{7}. \text{ global min.}$$

---

But except in simple examples, solving for $\nabla f(x) = 0$ just as difficult as opt. problem.

E.g. $\quad f(x) = x^2 + x + e^x \quad \leftarrow$ convex

$$\nabla f(x) = 2x + 1 + e^x, \quad \nabla^2 f(x) = 2 + e^x > 0, \ \forall x$$

Solving for $\nabla f(x) = 0$ needs iterative method.

## Example

$$f(x) = f(x_1, x_2) = 2x_1^2 + 0.5x_2^2 + 3x_1x_2 + 8x_1 + x_2 + 1$$

$$= \frac{1}{2} x^T \begin{bmatrix} 4 & 3 \\ 3 & 1 \end{bmatrix} x + [8 \; 1]\, x + 1$$

$$\nabla f(x) = \begin{bmatrix} 4 & 3 \\ 3 & 1 \end{bmatrix} x + \begin{bmatrix} 8 \\ 1 \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} 4 & 3 \\ 3 & 1 \end{bmatrix}$$ indefinite $\Rightarrow$ $f$ neither convex nor concave

$\nabla f(x) = 0$ has unique solution :

$x^* = (1, -4)$ stationary point

not extremum