Convex Functions cont’

Lecture 3 for 18660/18460: Optimization

Guannan Qu

January 20, 2026

Admin Stuff

- HW 1 will be released and due **Feb 5**
- Quiz for Lecture 3 out, due Jan 22 before lecture

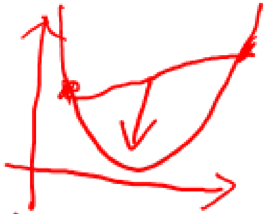
Recall

Typical functions and their convexity

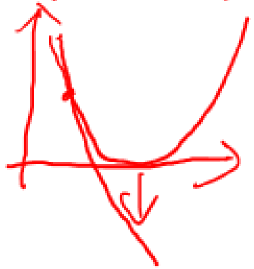
- Affine $f(x) = a^Tx + b$
- Quadratic $f(x) = x^TAx$ where A is symmetric & positive semi-definite
- Simple univariate functions e^x, x^a over $(0, +\infty)$ $(a \le 0, a \ge 1)$
  $-\log x$ over $(0, +\infty)$

Ways to check convexity

- Definition $\forall x, y$, $\forall t \in [0,1]$ $f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$



- First order and second order condition
  $\left\{ \begin{array}{l} \forall x, y, \ f(y) \ge f(x) + \nabla f(x)^T (y - x) \\ \forall x, \ \nabla^2 f(x) \succeq 0, \ \text{1-d case: } f''(x) \ge 0 \end{array} \right.$



Recall: Norm Functions

Recall: ℓ_p norms for a vector $x \in \mathbb{R}^n$ is defined by

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \quad \text{for } p \geq 1$$

Special cases:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \qquad \|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$$

Properties: Any norm $\|x\|$ satisfies

- For $a \in \mathbb{R}$, $\|ax\| = |a|\|x\|$
- $\|x\| = 0$ if and only if $x = 0$
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$

Recall: Norm Functions

Lemma. All norm functions are convex!

Proof: use triangle inequality

$$\|tx + (1-t)y\| \leq \|tx\| + \|(1-t)y\|$$

$$= t \cdot \|x\| + (1-t) \cdot \|y\|$$

Quiz Results

$\det|\lambda I - A|$

$|A - xI|$

Select all functions that are convex.

100 %	(A) $f(x) = e^x + 1$ over domain $\mathbb{R}$
12 %	(B) $f(x_1, x_2) = x_1 x_2$ over domain $\mathbb{R}^2$
88 %	(C) $f(x) = x\log x + (1-x)\log(1-x)$ over domain $(0, 1)$
84 %	(D) $f(x_1, x_2) = x_1^2 + x_2^2 + x_1 x_2$ over domain $\mathbb{R}^2$

(C) $f(x) = x\log x + (1-x)\log(1-x)$

$f'(x) = (x)'\cdot\log x + x\cdot(\log x)'$

$+ (1-x)'\log(1-x)$

$+ (1-x)(\log(1-x))'$

$= \log x + x\cdot\frac{1}{x} + (-1)\cdot\log(1-x)$

$+ (1-x)\cdot\frac{1}{1-x}\cdot(-1)$

$= \log x + 1 - \log(1-x) - 1$

$= \log x - \log(1-x)$

$f''(x) = \underbrace{\frac{1}{x}}_{>0} + \underbrace{\frac{1}{1-x}}_{>0} > 0$

(B) $\nabla^2 f(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

$\det\left|\lambda I - \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right| = \det\begin{vmatrix} \lambda - 0 & -1 \\ -1 & \lambda - 0 \end{vmatrix} = \lambda^2 - 1 = 0$

$\Rightarrow \begin{cases} \lambda = 1 \\ \lambda = -1 \end{cases} \Leftarrow$ means $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ NOT PSD.

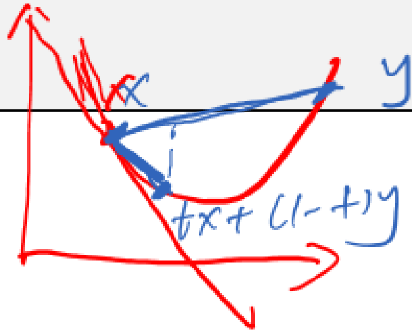
Tool 1: First Order Condition for Convexity

$$\lim_{t \to 0} \left[\frac{f(x + t(y-x)) - f(x)}{t(y-x)} \right] (y-x)$$

$$= f'(x)(y-x)$$

Lemma. A differentiable function $f(x)$ is convex if and only if for any x, y,

$$f(y) \geq f(x) + \nabla f(x)^{\top}(y - x). \quad (*)$$



How to prove this lemma? $\Rightarrow$, Assume f convex. prove $(*)$

By Def. of convex: $\forall\, x, y.\ t \in [0,1]$

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

"Directional Gradient"

$$\frac{f(x + t\cdot(y-x)) - f(x)}{t} \leq f(y) - f(x)$$

Def. of grad

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

let $t \to 0$

$$\nabla f(x)^{T}(y-x) \leq f(y) - f(x)$$

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

Tool 1: First Order Condition for Convexity

Lemma. A differentiable function $f(x)$ is convex if and only if for any x, y,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x). \quad (*)$$

How to prove this lemma? $\Leftarrow$ Assume $(*)$, prove convexity.

in other words, want to prove, $\forall x, y, t$

$$f((1-t)x + ty) \leq (1-t)f(x) + t f(y)$$

$$z = (1-t)x + ty$$

Apply $(*)$ to (x, z)

$$f(x) \geq f(z) + \nabla f(z)^T (x - z) \quad ①$$

Apply $(*)$ to (y, z)

$$f(y) \geq f(z) + \nabla f(z)^T (y - z) \quad ②$$

observation

$$\boxed{(1-t)(x-z) + t\cdot(y-z) = (1-t)x + ty - z = 0}$$

$① \times (1-t) + ② \times t$

$$(1-t)f(x) + t f(y)$$

$$\geq (1-t)f(z) + (1-t)\nabla f(z)^T (x-z) + t f(z) + t \nabla f(z)^T (y - z)$$

$$= f(z) + 0$$

Summary so far

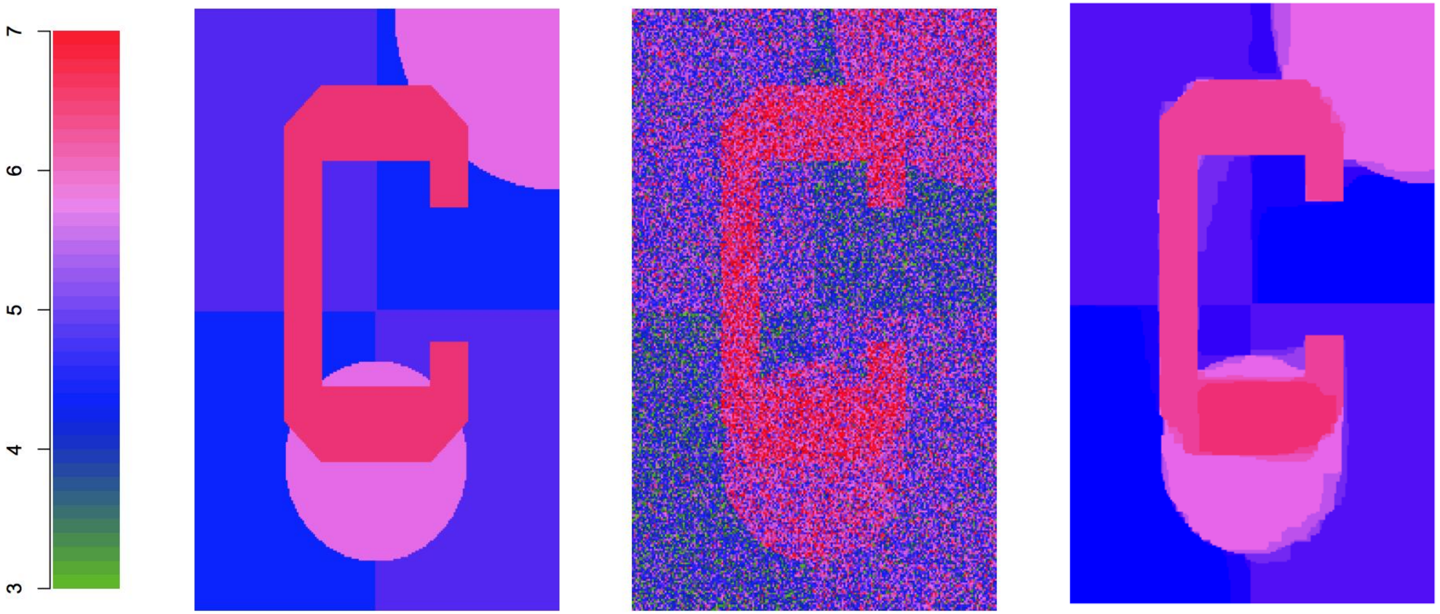
We have studied the convexity of

- Linear/affine functions
- Quadratic functions
- Simple univariate functions
- Norm functions

What about more sophisticated functions?

- e.g. $f(x) = x^2 + e^x$?
- e.g. $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$?
- e.g. the denoising example last lecture

Revisit: Denoising



True image　　Data　　Solution

$$\min_{\theta_1,\ldots,\theta_n} \underbrace{\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta_i)^2}_{\theta_i \text{ stays close to } y_i} + \lambda \underbrace{\sum_{(i,j)\ \text{adjacent}} |\theta_i-\theta_j|}_{\text{penalize changes in adjacent pixels}}$$

θ_i stays close to y_i　penalize changes in adjacent pixels

Revisit: Denoising

Is this function convex?

$$\min_{\theta_1,\ldots,\theta_n} \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta_i)^2+\lambda\sum_{(i,j)\ \mathrm{adjacent}}|\theta_i-\theta_j|$$

Summary so far

We have studied the convexity of

- Linear/affine functions
- Quadratic functions
- Simple univariate functions
- Norm functions

What about more sophisticated functions?

- e.g. $f(x) = x^2 + e^x$?
- e.g. $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$
- e.g. the denoising example last lecture
- Next: operations that preserve convexity!

Operations Preserving Convexity

Nonnegative linear combination

$f_1, \dots, f_m$ convex implies $\underbrace{a_1 f_1 + \dots + a_m f_m}_{f}$ convex for any $a_1, \dots a_m \geq 0$.

Proof: Checking Hessian

$$\nabla^2 f = \underbrace{a_1 \underbrace{\nabla^2 f_1}_{\text{psd}}}_{\text{psd}} + \underbrace{a_2 \underbrace{\nabla^2 f_2}_{\text{psd}}}_{\text{psd}} + \dots + \underbrace{a_m \underbrace{\nabla^2 f_m}_{\text{psd}}}_{\text{psd}} \quad \text{is p.s.d.}$$

Consequence: $f(x) = \underbrace{x^2}_{\text{psd}} + \underbrace{e^x}_{\text{psd}}$ is convex!

$f(x) = \underbrace{\frac{1}{2} x^\top A x + b^\top x + c}_{\text{psd}}$ is convex if A positive semi-definite!

Operations Preserving Convexity

If M is psd
Then $A^T M A$ is still psd

$m \times n$ $\quad$ $m \times 1$

Affine composition:
Suppose $h: \mathbb{R}^m \rightarrow \mathbb{R}$ is convex, then $f(x) = h(Ax + b)$ is convex where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

$$\underset{n\text{-d}}{x} \longrightarrow \underset{m\text{-d}}{Ax+b} \xrightarrow{h} \underset{1\text{-d}}{h(Ax+b)}$$

Proof $m = n = 1$ $\quad$ $A \in a$, $b \in b$

$$f(x) = h(ax+b)$$

$$f'(x) = \underline{h'(ax+b) \cdot a}$$

$$f''(x) = \underbrace{h''(ax+b)}_{\geq 0 \text{ because } h \text{ convex}} \cdot \underbrace{a \cdot a}_{\geq 0} \geq 0$$

$$f(x) = h(Ax+b)$$

$$\nabla f(x) = A^T \nabla h(Ax+b)$$

$$\nabla^2 f(x) = \underbrace{A^T \underbrace{\nabla^2 h(Ax+b)}_{\text{psd}} \cdot A}_{\text{psd.}}$$

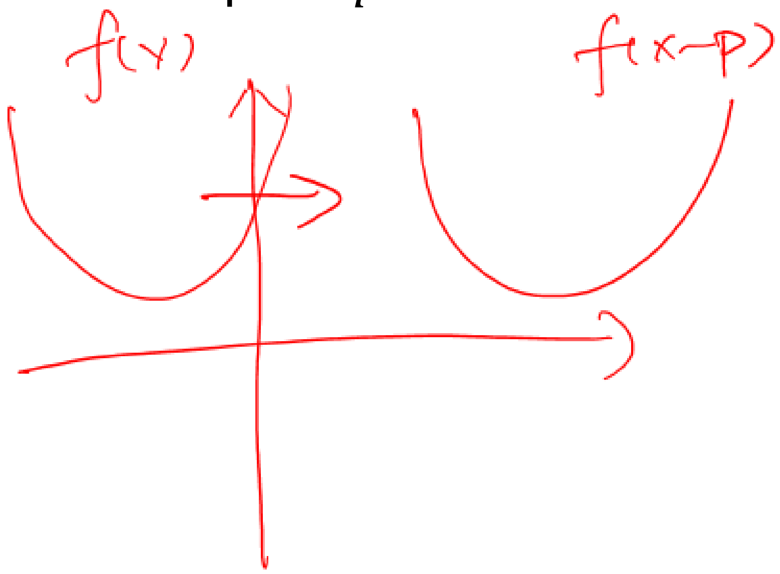
Operations Preserving Convexity

Affine composition:
Suppose $h: \mathbb{R}^m \to \mathbb{R}$ is convex, then $f(x) = h(Ax + b)$ is convex where $\mathrm{A} \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^n$.

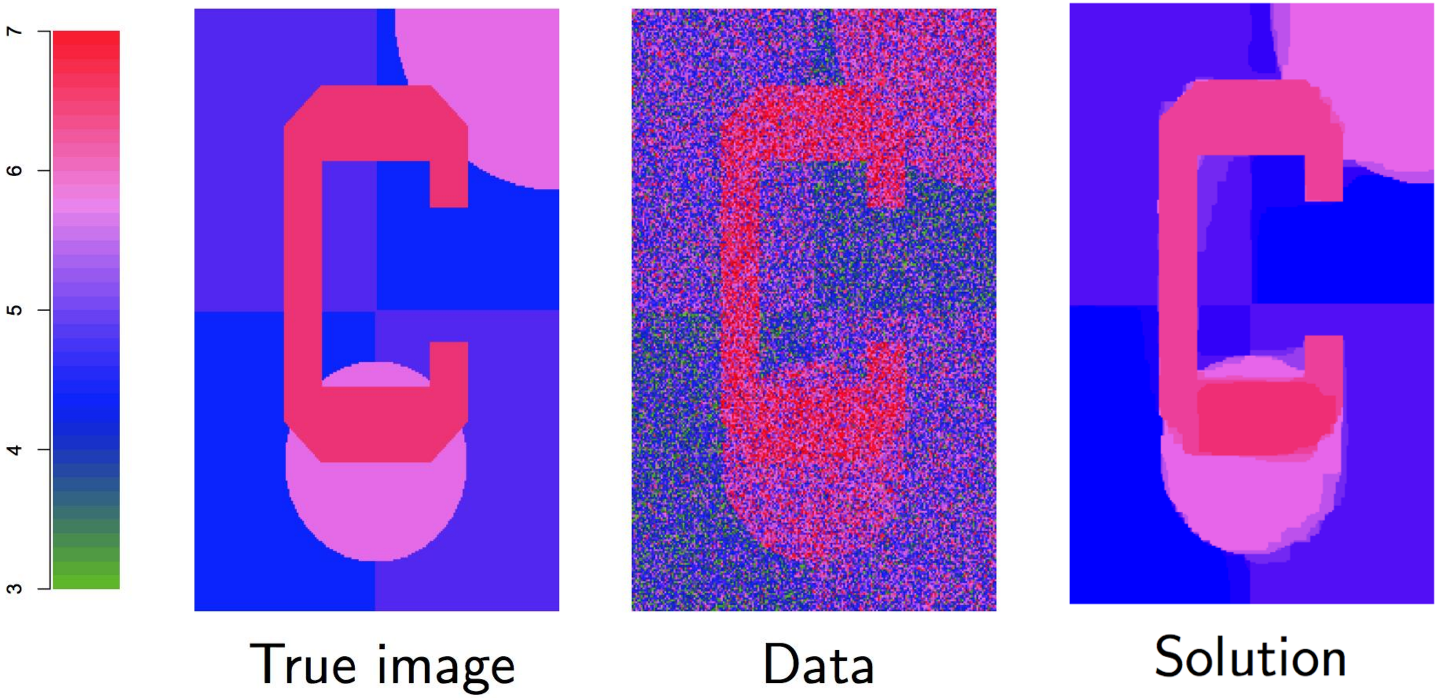
Consequence

$A = I \quad b = -p$

- Given point $p \in \mathbb{R}^n$ and convex function $f(x)$, $f(x - p)$ is convex.



Revisit: Denoising



$$\min_{\theta_1,\ldots,\theta_n} \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta_i)^2 + \lambda \sum_{(i,j)\ \text{adjacent}} |\theta_i-\theta_j|$$

θ_i stays close to y_i penalize changes in adjacent pixels

Revisit: Denoising

$[\nabla^2 g(\theta)]_{l,k} \neq 0$ if $l,k = i$

$$\min_{\theta_1,\dots,\theta_n} \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta_i)^2 + \lambda \sum_{(i,j)\text{ adjacent}} |\theta_i - \theta_j|$$

$\lambda > 0$

Show that this is a convex function:

$g(\theta) = (y_i - \theta_i)^2 = \underbrace{\theta_i^2}_{\text{convex}} - \underbrace{2y_i\theta_i + y_i^2}_{\text{affine in } \theta_i}$ convex

convex

$[0 \cdots, 0, 1, \cdots]^T \begin{bmatrix}\theta_1 \\ \vdots \\ \theta_i \\ \vdots \\ \theta_n\end{bmatrix}$

$\rightarrow |\theta_i - \theta_j| = |\underbrace{[1 \; -1]\begin{bmatrix}\theta_i \\ \theta_j\end{bmatrix}}_{\theta_i - \theta_j}| = \underbrace{h}_{\text{convex}}(\underbrace{[1 \; -1]\begin{bmatrix}\theta_i \\ \theta_j\end{bmatrix}}_{\text{affine transformation}}) \Rightarrow$ convex

$h(x) = |x|$

Operations Preserving Convexity

General composition:
Suppose $f(x) = h(g(x))$, where $g : \mathbb{R}^n \to \mathbb{R}, h: \mathbb{R} \to \mathbb{R}, \; f: \mathbb{R}^n \to \mathbb{R}$. Then when

- g is convex
- h is nondecreasing and convex

$h' \geq 0$

Then $f(x) = h(g(x))$ is convex.

Example: why the "nondecreasing" condition is necessary

$g(x) = x^2$ convex

$h(y) = -y$ convex decreasing

$h(g(x)) = -x^2$ NOT convex

Operations Preserving Convexity

General composition:
Suppose $f(x) = h(g(x))$, where $g : \mathbb{R}^n \to \mathbb{R}, h: \mathbb{R} \to \mathbb{R}$, $f: \mathbb{R}^n \to \mathbb{R}$. Then when

- g is convex
- h is nondecreasing and convex

Then $f(x) = h(g(x))$ is convex.

Example: logistic regression $f(x) = -\log \frac{1}{1+e^{a^\top x}}$

Operations Preserving Convexity

General composition:
Suppose $f(x) = h(g(x))$, where $g : \mathbb{R}^n \to \mathbb{R}, h: \mathbb{R} \to \mathbb{R}, f: \mathbb{R}^n \to \mathbb{R}$. Then when

- g is convex
- h is nondecreasing and convex

Then $f(x) = h(g(x))$ is convex.

Proof: x 1-d :

$$f'(x) = h'(g(x)) \cdot g'(x)$$

$$f''(x) = [h(g(x))]' g'(x) + h'(g(x)) [g'(x)]'$$

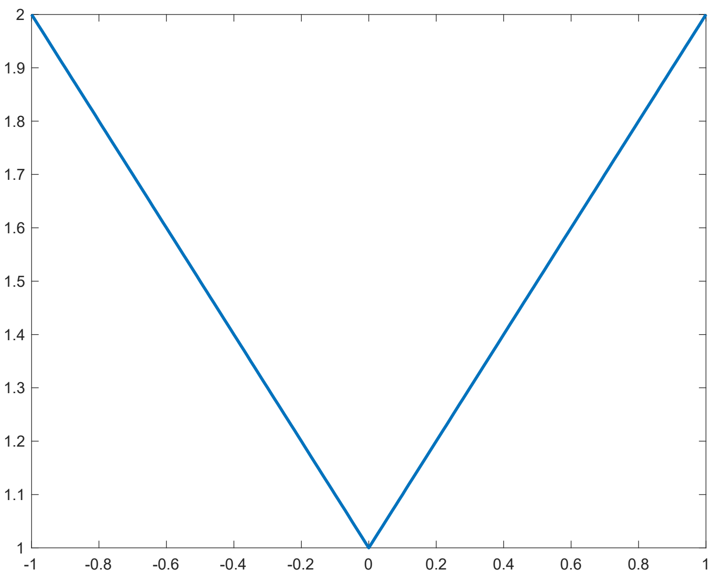
$$= \underbrace{h''(g(x))}_{\geq 0} \underbrace{g'(x) \cdot g'(x)}_{\geq 0} + \underbrace{h'(g(x))}_{\geq 0} \underbrace{g''(x)}_{\geq 0} \geq 0$$

non-decreasing of h

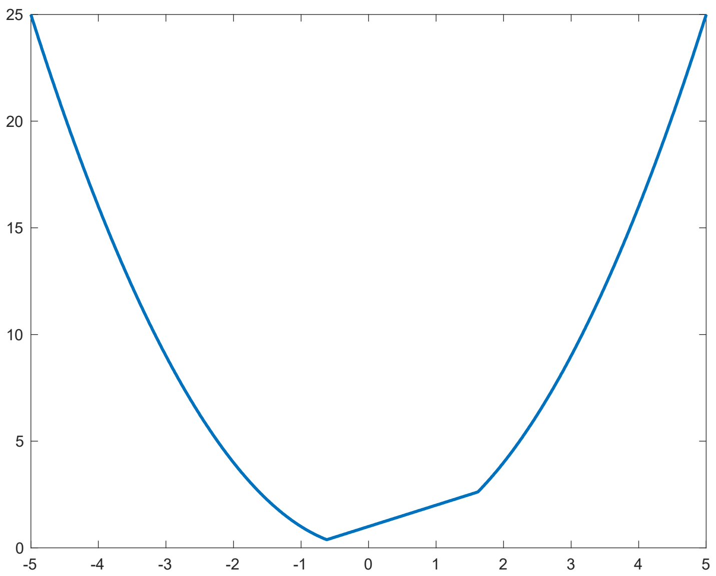
Operations Preserving Convexity

Pointwise maximization:
If $f_1, f_2, \dots, f_m$ is convex, then $f(x) = \max(f_1(x), \dots, f_m(x))$ is convex.

$f(x) = \max(x+1, -x+1)$



$f(x) = \max(x+1, x^2)$



Operations Preserving Convexity

Pointwise maximization:
If $f_1, f_2, \dots, f_m$ is convex, then $f(x) = \max(f_1(x), \dots, f_m(x))$ is convex.

Proof: we use definition. Given $t \in [0,1]$ and x, y,

Operations Preserving Convexity

Pointwise maximization:

If f_i is convex for all $i \in I$ where I may be an infinite set, then

$$f(x) = \max_{i \in I} f_i(x)$$

is convex.

Proof: identical

Summary: Operations Preserving Convexity

- Nonnegative linear combination

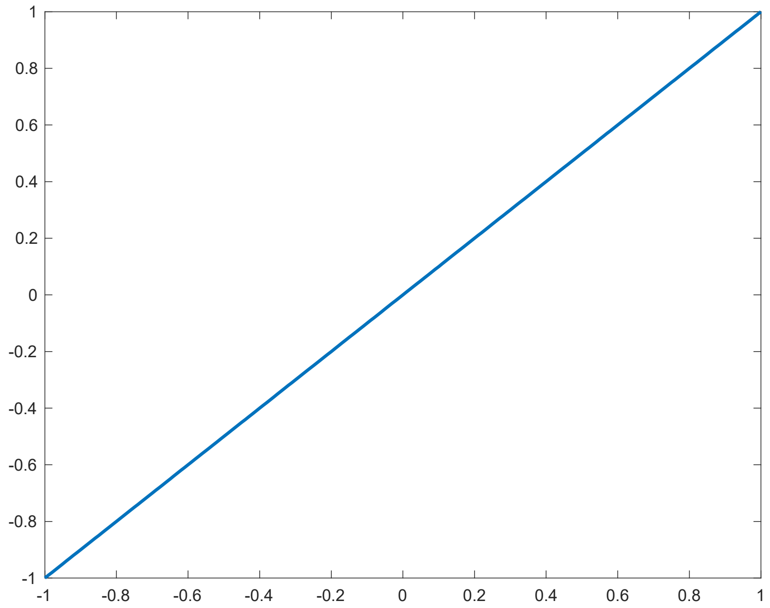
- Compositions

- Pointwise maximization

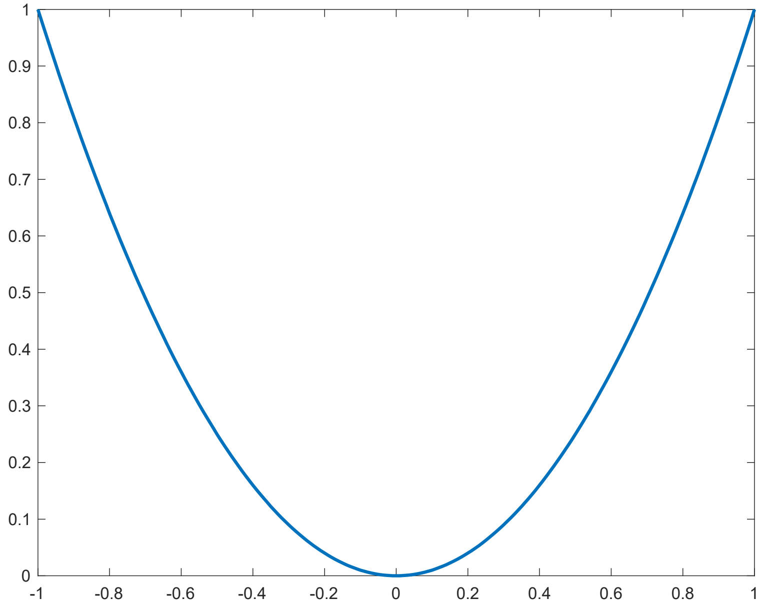
Next: strict convexity, strong convexity, smoothness

Strict Convexity and Strong Convexity

Both these functions are convex, but they look different



An affine function



A quadratic function

Strict Convexity

Definition. A function $f(x)$ is strictly convex if for any $x \neq y$, $t \in (0,1)$

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y).$$

Equivalent definition using first order condition:

$$f(y) > f(x) + \nabla f(x)^{\top}(y - x).$$

Equivalent definition using second order condition:

$$\nabla^2 f(y) \succ 0.$$

Strong Convexity

Definition. For $\mu > 0$, a function $f(x)$ is μ-strongly convex if $f(x) - \frac{\mu}{2}\|x\|^2$ is convex

Equivalent definition using first order condition: for any x, y

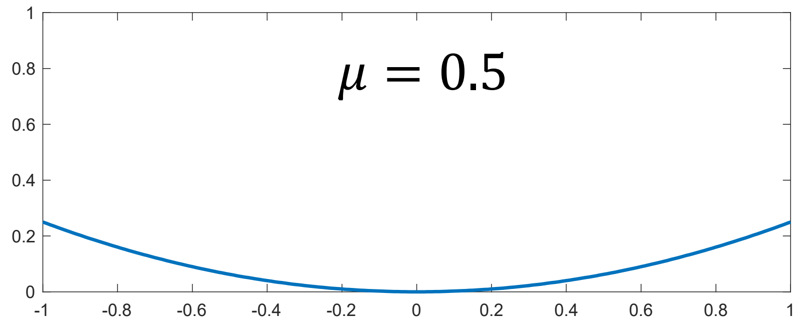
$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2}\|y - x\|^2.$$

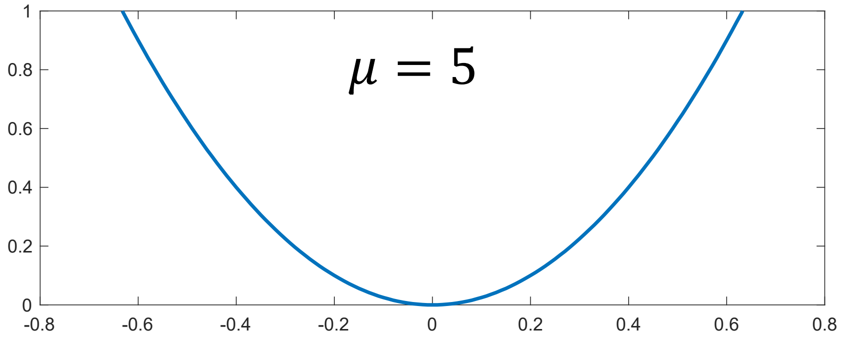
Equivalent definition using second order condition: for any x

$$\nabla^2 f(x) \succcurlyeq \mu I$$

How to understand strong convexity?

- Use quadratic function $\frac{\mu}{2}\|x\|^2 = \frac{\mu}{2} x^\top x$ as a reference to tell how "curved" functions are





Strong Convexity

Definition. For $\mu > 0$, a function $f(x)$ is μ-strongly convex if $f(x) - \frac{\mu}{2}\|x\|^2$ is convex

Equivalent definition using first order condition: for any x, y

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2}\|y - x\|^2.$$

Equivalent definition using second order condition: for any x

$$\nabla^2 f(x) \succcurlyeq \mu I$$

How to understand strong convexity?

- Use quadratic function $\frac{\mu}{2}\|x\|^2 = \frac{\mu}{2}x^\top x$ as a reference to tell how "curved" functions are
- Why choosing $\frac{\mu}{2}\|x\|^2$ as the reference? - its Hessian is μI, a simple matrix

More on Strong Convexity

Definition. For $\mu > 0$, a function $f(x)$ is μ-strongly convex if $f(x) - \frac{\mu}{2}\|x\|^2$ is convex

Equivalent definition using first order condition: for any x, y

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2}\|y - x\|^2.$$

Equivalent definition using second order condition: for any x

$$\nabla^2 f(x) \succcurlyeq \mu I$$

Strong convexity of general quadratic functions $f(x) = \frac{1}{2}x^\top A x$

$$\nabla^2 f(x) = A \succcurlyeq \sigma_{min}(A) I$$

When A is positive definite (i.e. $\sigma_{min}(A) > 0$), $f(x) = \frac{1}{2}x^\top A x$ is $\sigma_{min}(A)$ -strongly convex

Comparison btw Strict/Strong convexity

strict convexity

$$\nabla^2 f(x) \succ 0, \forall x \in dom(f)$$

Example: $f(x) = \frac{1}{x}$ over $(0, +\infty)$

μ-strong convexity

$$\nabla^2 f(x) \succcurlyeq \mu I, \forall x \in dom(f)$$

Example: $f(x) = \frac{1}{2} a x^2$ over $\mathbb{R}$, $a > 0$

Concavity

A function f is concave if $-f$ is convex
A function f is strictly concave if $-f$ is strictly convex
A function f is μ-strongly concave if $-f$ is μ-strongly convex

	Definition	First-order condition	Second-order condition
concave	$f(tx+(1-t)y) \geq tf(x)+(1-t)f(y)$, $\forall x, y, \forall t \in [0,1]$	$f(y) \leq f(x) + \nabla f(x)^{\top}(y-x), \forall x, y$	$\nabla^2 f(x) \preccurlyeq 0$, $\forall x$
convex	$f(tx+(1-t)y) \leq tf(x)+(1-t)f(y)$, $\forall x, y, \forall t \in [0,1]$	$f(y) \geq f(x) + \nabla f(x)^{\top}(y-x), \forall x, y$	$\nabla^2 f(x) \succcurlyeq 0$, $\forall x$
strictly concave	$f(tx+(1-t)y) > tf(x)+(1-t)f(y)$, $\forall x \neq y, \forall t \in (0,1)$	$f(y) < f(x) + \nabla f(x)^{\top}(y-x), \forall x \neq y$	$\nabla^2 f(x) \prec 0$, $\forall x$
strictly convex	$f(tx+(1-t)y) < tf(x)+(1-t)f(y)$, $\forall x \neq y, \forall t \in (0,1)$	$f(y) > f(x) + \nabla f(x)^{\top}(y-x), \forall x \neq y$	$\nabla^2 f(x) \succ 0$, $\forall x$
μ-strongly concave	$f(x) + \frac{\mu}{2}\|x\|^2$ is concave	$f(y) \leq f(x) + \nabla f(x)^{\top}(y-x) - \frac{\mu}{2}\|y-x\|^2$, $\forall x, y$	$\nabla^2 f(x) \preccurlyeq -\mu I$, $\forall x$
μ-strongly convex	$f(x) - \frac{\mu}{2}\|x\|^2$ is convex	$f(y) \geq f(x) + \nabla f(x)^{\top}(y-x) + \frac{\mu}{2}\|y-x\|^2$, $\forall x, y$	$\nabla^2 f(x) \succcurlyeq \mu I$, $\forall x$

Summary for Convex Functions

Typical functions and their convexity

- Affine, quadratic
- Simple univariate functions
- Norm functions

Ways to check convexity

- Definition
- First order and second order condition
- Operations that preserve convexity

Strict and strong convexity

Next: smoothness, convex constraint sets