Optimization in Machine Learning

Lecture 5: Convex Functions, Strong Convexity, Calculus of Convexity, ML Examples

Ganesh Ramakrishnan

Department of Computer Science
Dept of CSE, IIT Bombay
https://www.cse.iitb.ac.in/~ganesh

January, 2025

Outline of Content for Today

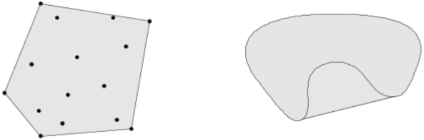
- Convex Sets and Seperating/Supporting Hyperplane Theorem
  1. Definition, Properties and Examples of Convex Sets [Done]
  2. Discussion on Homework problem
- Convexity of Functions
  1. Definition of Convexity, Strong Convexity and Strict Convexity
  2. Examples of Convex Functions
  3. Calculus of Convex Functions & More Properties
  4. Basic Subgradient Calculus: Subgradients for non-differentiable convex functions
  5. Understanding the Convexity of Machine Learning Loss Functions
  6. Direction Vector, Subgradients and Subdifferentials, Epigraphs and Sublevel sets,
  7. First Order Convexity Conditions, Quasi Convexity
  8. Calculus of Subgradients,
  9. Convex Optimization Problems
  10.

[Recap] Convex combination and convex hull

- **Convex combination** of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ is any point $\mathbf{x}$ of the form

$$\begin{aligned} \mathbf{x} = \quad & \theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 + ... + \theta_k \mathbf{x}_k = conv(\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\}) \\ \text{with} \quad & \theta_1 + \theta_2 + ... + \theta_k = 1, \theta_i \geq 0. \end{aligned}$$

- **Convex hull or conv(S)** is the set of all convex combinations of point in the set S.



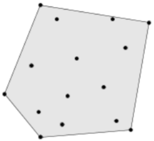
- Should S be always convex?
- What about the convexity of conv(S)?

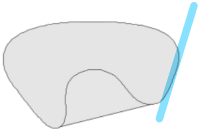
[Recap] Convex combination and convex hull

- **Convex combination** of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ is any point $\mathbf{x}$ of the form

$$\begin{aligned} \mathbf{x} = &\quad \theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 + ... + \theta_k \mathbf{x}_k = conv(\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\}) \\ \text{with} &\quad \theta_1 + \theta_2 + ... + \theta_k = 1, \theta_i \geq 0. \end{aligned}$$

- **Convex hull or conv(S)** is the set of all convex combinations of point in the set S.





- Should S be always convex?
- What about the convexity of conv(S)?

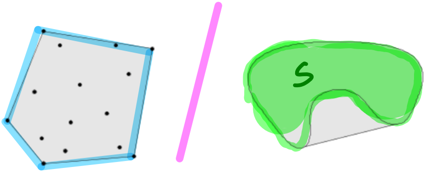
H/W: Is there a notion of "Supporting hyperplane" that are characteristic to convex sets and not found in non-convex sets?

[Recap] Convex combination and convex hull

- **Convex combination** of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ is any point $\mathbf{x}$ of the form

$$\begin{aligned} \mathbf{x} = & \quad \theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 + ... + \theta_k \mathbf{x}_k = conv(\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\}) \\ \text{with} & \quad \theta_1 + \theta_2 + ... + \theta_k = 1, \theta_i \geq 0. \end{aligned}$$

- **Convex hull or conv(S)** is the set of all convex combinations of point in the set S.



Homework: S is given to you! The S above is connected but not convex. Hence no need for S to be convex

- Should S be always convex? No.
- What about the convexity of conv(S)? It's always convex.

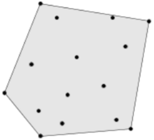
[Recap] Convex combination and convex hull

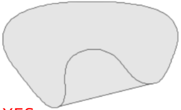
- **Convex combination** of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ is any point $\mathbf{x}$ of the form

$$\mathbf{x} = \theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 + ... + \theta_k\mathbf{x}_k = conv(\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\})$$

$$\text{with} \quad \theta_1 + \theta_2 + ... + \theta_k = 1, \theta_i \geq 0.$$

- **Convex hull or conv(S)** is the set of all convex combinations of point in the set S.





YES

Verify that by substituting for x_1 and x_2 in terms of y_1, y_2, z_1 and z_2 the new combination is still convex in y_1, y_2, z_1 and z_2

MANDATORY HOMEWORK: Please verify! $\lambda_1 x_1 + (1-\lambda_1) x_2$

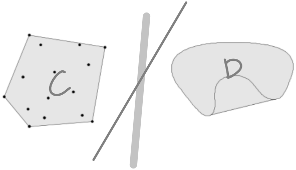
$\theta_1 y_1 + (1-\theta_1) y_2$ $\quad \theta_2 z_1 + (1-\theta_2) z_2$

Assume: $y_1, y_2, z_1, z_2 \in S$ & $x_1, x_2 \in$ conv(S)

- Should S be always convex? No.
- What about the convexity of conv(S)? It's always convex.

SHT: Separating hyperplane theorem (a fundamental theorem) [Homework solution]

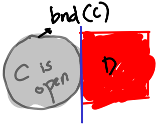
If $\mathcal{C}$ and $\mathcal{D}$ are disjoint convex sets, *i.e.*, $\mathcal{C} \cap \mathcal{D} = \phi$, then there exists $\mathbf{a} \neq \mathbf{0}$, with a $b \in \Re$ such that
$\mathbf{a}^T\mathbf{x} \leq \mathbf{b}$ for $\mathbf{x} \in \mathcal{C}$,
$\mathbf{a}^T\mathbf{x} \geq \mathbf{b}$ for $\mathbf{x} \in \mathcal{D}$.
That is, the hyperplane $\left\{\mathbf{x} | \mathbf{a}^T\mathbf{x} = \mathbf{b}\right\}$ separates $\mathcal{C}$ and $\mathcal{D}$.



This is an example of strict separation

- The seperating hyperplane need not be unique though.
- Strict separation requires additional assumptions (e.g., C is closed, D is a singleton).

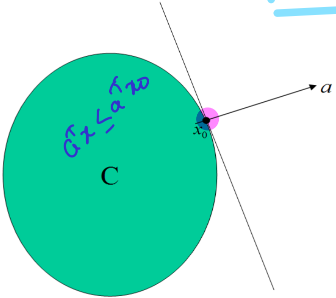
C is open
if bnd(C) $\notin$ C



Supporting hyperplane theorem (consequence of separating hyperplane theorem) [Homework solution]

Supporting hyperplane to set $\mathcal{C}$ at boundary point $\mathbf{x}_o$:

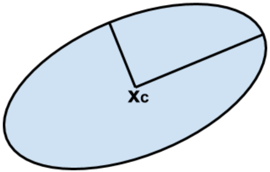
- $\{\mathbf{x} | \mathbf{a}^T\mathbf{x} = \mathbf{a}^T\mathbf{x}_o\}$
- where $\mathbf{a} \neq 0$ and $\boxed{\mathbf{a}^T\mathbf{x} \leq \mathbf{a}^T\mathbf{x}_o}$ for all $\mathbf{x} \in \mathcal{C}$



Supporting hyperplane theorem: if $\mathcal{C}$ is convex, then there exists a supporting hyperplane at every boundary point of C.

Euclidean balls and ellipsoids

- **Euclidean ball** with **center** $\mathbf{x}_c$ and **radius** r is given by:
  $B(\mathbf{x}_c, r) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\|_2 \le r\} = \{\mathbf{x}_c + r\mathbf{u} \mid \|\mathbf{u}\|_2 \le 1\}$
- **Ellipsoid** is a **set** of form:
  $\left\{\mathbf{x} \,\middle|\, (\mathbf{x} - \mathbf{x}_c)^T P^{-1}(\mathbf{x} - \mathbf{x}_c) \le 1\right\}$, where $P \in S^n_{++}$ i.e. P is SPD matrix.
  - Other representation: $\{\mathbf{x}_c + A\mathbf{u} \mid \|\mathbf{u}\|_2 \le 1\}$ s.t. A is square & non-singular (*i.e.*, A^{-1} exists).
  - It turns out that $A = P^{\frac{1}{2}}$, which you can easily verify by substituting.

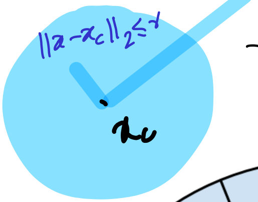


- **Euclidean ball** with **center** $\mathbf{x}_c$ and **radius** r is given by:
  $B(\mathbf{x}_c, r) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\|_2 \le r\} = \{\mathbf{x}_c + r\mathbf{u} \mid \|\mathbf{u}\|_2 \le 1\}$
- **Ellipsoid** is a **set** of form:
  $\{\mathbf{x} \mid (\mathbf{x} - \mathbf{x}_c)^T P^{-1} (\mathbf{x} - \mathbf{x}_c) \le 1\}$, where $P \in S^n_{++}$ i.e. P is SPD matrix.
  - Other representation: $\{\mathbf{x}_c + A\mathbf{u} \mid \|\mathbf{u}\|_2 \le 1\}$ s.t. A is square & non-singular (*i.e.*, A^{-1} exists).
  - It turns out that $A = P^{\frac{1}{2}}$, which you can easily verify by substituting.

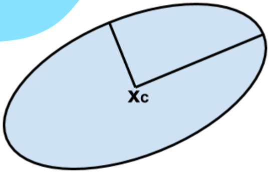
Symmetric & PD

$[x^T P x > 0]$

[Eigenvalues > 0]



P is bringing about non-uniform scaling+rotation



S^n_{++}

A space of nxn symmetric matrices that are positive definite

Norm balls

- **Recap Norm:** A function[1] $\|.\|$ that satisfies:
  1. $\|\mathbf{x}\| \geq 0$, and $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$.
  2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for any scalar $\alpha \in \Re$.
  3. $\|\mathbf{x}_1 + \mathbf{x}_2\| \leq \|\mathbf{x}_1\| + \|\mathbf{x}_2\|$ for any vectors $\mathbf{x}_1$ and $\mathbf{x}_2$.
- **Norm ball** with **center** $\mathbf{x}_c$ and **radius** r: $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}_c\| \leq r\}$ is a convex set. Why?

[1] ($\|.\|$ is a general (unspecified) norm; $\|.\|_{symb}$ is particular norm.)

Norm balls

To prove:1 <= 6
Follows from assumption: 4 <= 5 $(x_1, x_2 \in C)$
2 =3 (green become pink and yellow become gray)
3 <= 4 by the triangle inequality for norms

$$\|p + q\| \leq \|p\| + \|q\|$$

Hint: Use this triangle inequality

- **Recap Norm:** A function[1] $\|.\|$ that satisfies:
  1. $\|\mathbf{x}\| \geq 0$, and $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$.
  2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for any scalar $\alpha \in \Re$.
  3. $\|\mathbf{x}_1 + \mathbf{x}_2\| \leq \|\mathbf{x}_1\| + \|\mathbf{x}_2\|$ for any vectors $\mathbf{x}_1$ and $\mathbf{x}_2$.
- B **Norm ball** with **center** $\mathbf{x}_c$ and **radius** r: $\{\mathbf{x} \,|\, \boxed{\|\mathbf{x} - \mathbf{x}_c\| \leq r}\}$ is a convex set. Why?

Ans: For any $x_1, x_2 \in B$

$$\underbrace{\|(\lambda x_1 + (1-\lambda)x_2) - x_c\|_p}_{\text{To prove}} \overset{(1)}{=} \|\lambda x_1 + (1-\lambda)x_2 - \lambda x_c - (1-\lambda)x_c\|_p \; (2)$$

$$= \|\lambda(x_1 - x_c) + (1-\lambda)(x_2 - x_c)\|_p \; (3) \leq \lambda\|x_1 - x_c\| + (1-\lambda)\|x_2 - x_c\| \; (4)$$

$$\leq \lambda r + (1-\lambda) r \; (5) = \underbrace{r}_{\text{To prove}} \; (6)$$

[1] ($\|.\|$ is a general (unspecified) norm; $\|.\|_{symb}$ is particular norm.)

Norm balls

- **Recap Norm:** A function[1] $\|.\|$ that satisfies:
  1. $\|\mathbf{x}\| \geq 0$, and $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$.
  2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for any scalar $\alpha \in \Re$.
  3. $\|\mathbf{x}_1 + \mathbf{x}_2\| \leq \|\mathbf{x}_1\| + \|\mathbf{x}_2\|$ for any vectors $\mathbf{x}_1$ and $\mathbf{x}_2$.
- **Norm ball** with **center** $\mathbf{x}_c$ and **radius** r: $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}_c\| \leq r\}$ is a convex set. Why?
  - Eg 1: **Ellipsoid** is defined using $\|\mathbf{x}\|_P^2 = \mathbf{x}^T P\mathbf{x}$.
  - Eg 2: **Euclidean ball** is defined using $\|\mathbf{x}\|_2$.

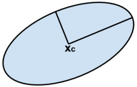
[1] ($\|.\|$ is a general (unspecified) norm; $\|.\|_{symb}$ is particular norm.)

Solution to the Problems

- Every Norm is a Convex function. **Why?**
  - By Triangle Inequality: $f(x+y) \leq f(x) + f(y)$, and homogeneity of norm: $f(\alpha x) = \alpha f(x)$ for a scalar α
  - It follows that

$$f(\lambda x + (1-\lambda)y) \leq f(\lambda x) + f((1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y)$$

- **Ellipsoid** is a **convex set** with two equivalent forms:
  $\left\{\mathbf{x} \mid (\mathbf{x}-\mathbf{x}_c)^T P^{-1}(\mathbf{x}-\mathbf{x}_c) \leq 1\right\}$, where $P \in S^n_{++}$ i.e. P is SPD matrix.
  OR $\{\mathbf{x}_c + A\mathbf{u} \mid \|\mathbf{u}\|_2 \leq 1\}$ s.t. A is square & non-singular ($i.e.$, A^{-1} exists).
  It turns out that $A = P^{\frac{1}{2}}$, which you can easily verify by substituting.



Convex Functions

- A Function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is convex if:
  - $dom(f)$ is a convex set
  - for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
- Geometrically, the line segment between $(x, f(x))$ and $(y, f(y))$ lies above the graph of f.

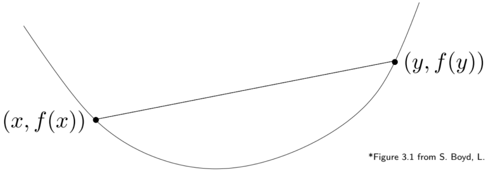


*Figure 3.1 from S. Boyd, L. Vandenberghe

- f is strictly convex if for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$

Convex Functions

The second interesting set shown on the board
| x | f(x) <= z} for some fixed z

Defined by z

Defined by z

- A Function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is convex if:
  - $dom(f)$ is a convex set
  - Function at convex combination <= convex combination of the fn values
  - for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
- Geometrically, the line segment between $(x, f(x))$ and $(y, f(y))$ lies above the graph of f.

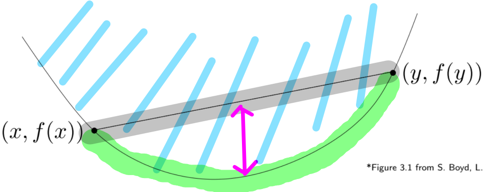


*Figure 3.1 from S. Boyd, L. Vandenberghe

The first interesting set shown on the board:
{(x,z) | f(x) <=z}

- f is strictly convex if for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have:
$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$

Strongly Convex Functions

- A Function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is strongly convex if there exists a $\mu > 0$ such that the function $g(x) = f(x) - \mu/2\|x\|^2$ is convex
- The parameter μ is the strong convexity parameter
- Geometrically, strong convexity means that there exists a quadratic lower bound on the growth of the function.
- Its easy to see that Strong Convexity implies Strict Convexity!
- Strong Convexity Doesn't imply the function is differentiable!
- If a function f is strongly convex and g is convex (not necessarily strongly convex), $f + g$ is strongly convex.
- $\|x\|^2$ is strongly convex!
- Hence for any convex function f, the function $f(x) + \lambda/2\|x\|^2$ is strongly convex!

Strongly Convex Functions

Equality when $\theta = 0$ or 1 or $x_1 = x_2$

HOMEWORK1: Derive the term T such that $f(\theta x_1 + (1-\theta)x_2) + T \leq \theta f(x_1) + (1-\theta) f(x_2)$

in terms of d

- A Function $f : \mathbb{R}^d \mapsto \mathbb{R}$ is strongly convex if there exists a $\mu > 0$ such that the function $g(x) = f(x) - \mu/2\|x\|^2$ is convex
- The parameter μ is the strong convexity parameter
- Geometrically, strong convexity means that there exists a quadratic lower bound on the growth of the function.
- Its easy to see that Strong Convexity implies Strict Convexity!
- Strong Convexity Doesn't imply the function is differentiable! Homework2 using the hint provided below
- If a function f is strongly convex and g is convex (not necessarily strongly convex), $f + g$ is strongly convex. Homework3
- $\|x\|^2$ is strongly convex!
- Hence for any convex function f, the function $f(x) + \lambda/2\|x\|^2$ is strongly convex!

Solution to the Problem: Strongly Convex Functions

Prove that:
A function $f : \mathbb{R}^d \to \mathbb{R}$ is strongly convex if there exists a $\mu > 0$ such that the function $g(x) = f(x) - \mu/2\|x\|^2$ is convex
$\Rightarrow$ function $f : \mathbb{R}^d \to \mathbb{R}$ is strongly convex if for all $\mathbf{x}_1, \mathbf{x}_2 \in dom(f)$ and $\theta : 0 < \theta < 1$, we have: $f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq \theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2) - \frac{1}{2}\mu\theta(1-\theta)\|\mathbf{x}_1 - \mathbf{x}_2\|^2$
And vice-versa.
Hence for any convex function f, the function $f(x) + \lambda/2\|x\|^2$ is strongly convex!

Convex Functions

Examples of Convex Functions

- Linear Functions: $f(x) = a^T x$
- Affine Functions: $f(x) = a^T x + b$
- Exponential: $f(x) = exp(\alpha x)$
- Every Norm is Convex. **Why?**

RECAP HOW WE PROVED THAT NORM BALLS ARE CONVEX SETS

Ans: For any $x_1, x_2 \in B$

$$\|(\lambda x_1 + (1-\lambda)x_2) - x_c\|_p = \|\lambda x_1 + (1-\lambda)x_2 - \lambda x_c - (1-\lambda)x_c\|_p$$

$$= \|\lambda(x_1 - x_c) + (1-\lambda)(x_2 - x_c)\|_p \leq \lambda\|x_1 - x_c\| + (1-\lambda)(x_2 - x_c)$$

Setting x_c = 0
and replace lambda with theta
...we get

$$\|\theta x_1 + (1-\theta)x_2\| \leq \theta\|x_1\| + (1-\theta)\|x_2\|$$

By similar application of triangle inequality

Examples of Convex Functions

- Linear Functions: $f(x) = a^T x$
- Affine Functions: $f(x) = a^T x + b$
- Exponential: $f(x) = exp(\alpha x)$
- Every Norm is Convex. **Why?**
    - By Triangle Inequality: $f(x + y) \leq f(x) + f(y)$, and homogeneity of norm: $f(\alpha x) = \alpha f(x)$ for a scalar α

Convex Functions

Examples of Convex Functions

- Linear Functions: $f(x) = a^T x$
- Affine Functions: $f(x) = a^T x + b$
- Exponential: $f(x) = exp(\alpha x)$
- Every Norm is Convex. **Why?**
  - By Triangle Inequality: $f(x+y) \leq f(x) + f(y)$, and homogeneity of norm: $f(\alpha x) = \alpha f(x)$ for a scalar α
  - It follows that

$$f(\lambda x + (1-\lambda)y) \leq f(\lambda x) + f((1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y)$$

Properties of Convex Functions

- **Non-negative weighted sum:** $f = \sum_{i=1}^{n} \alpha_i f_i$ is convex if each f_i for $1 \leq i \leq n$ is convex and $\alpha_i \geq 0, 1 \leq i \leq n$.
- **Composition with Affine function:** $f(Ax + b)$ is convex if f is convex. For example:
  - The log barrier for linear inequalities, $f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$, is convex since $-\log(x)$ is convex.
  - Any norm of an affine function, $f(x) = ||Ax + b||$, is convex.

Composition with Scalar Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$.

$$f(x) = h(g(x))$$

- f is convex if a) g convex, h convex and non-decreasing or b) g concave, h convex and non-increasing
- Proof idea: First principles (definition of convexity in terms of pairs of points).

Composition with Scalar Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$.

$$f(x) = h(g(x))$$

- f is convex if a) g convex, h convex and non-decreasing or b) g concave, h convex and non-increasing
- Proof idea: First principles (definition of convexity in terms of pairs of points).

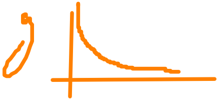
Composition with Scalar Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$.

$$f(x) = h(g(x))$$

- f is convex if a) g convex, h convex and non-decreasing or b) g concave, h convex and non-increasing
- Proof idea: First principles (definition of convexity in terms of pairs of points). (Advanced) For doubly differentiable functions: Take double derivative and try to show that $\nabla^2 f \geq 0$ (easier to prove this for $m = 1$).
- Examples:
  - $\exp(f(x))$ is convex if f is convex
  - $1/g(x)$ is convex if g is concave and positive

H/w: What if
g is allowed
to be negative?



Composition with Vector Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$.

$$f(x) = h(g(x)) = h\left(g_1(x), \cdots, g_k(x)\right)$$

- f is convex if a) g_i's convex, h convex and non-decreasing in each argument or b) g_i concave, h convex and non-increasing in each argument
- Examples:
  - $f(x) = \sum_i \log(g_i(x))$ is concave if g_i is concave and positive
  - $\log \sum_{i=1}^k \exp(g_i(x))$ is convex if g_i is convex.

Solution to Problem: Composition with Vector Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$.

$$f(x) = h(g(x)) = h\left(g_1(x), \cdots, g_k(x)\right)$$

- f is convex if a) g_i's are convex, h convex and non-decreasing in each argument or b) g_i's are concave, h convex and non-increasing in each argument
- Examples:
  - $f(x) = \sum_i \log(g_i(x))$ is concave if g_i is concave and positive
  - $\log \sum_{i=1}^{k} \exp(g_i(x))$ is convex if g_i is convex. Hints?
    - ★ A function r is convex iff $r\left(\frac{x_1+x_2}{2}\right) \leq \frac{1}{2}r(x_1) + \frac{1}{2}r(x_2)$ (special case with $\alpha = \frac{1}{2}$ is equivalent to to the general case with $\alpha \in [0, 1]$)

NOTE: FROM THE PROOF BELOW, IT APPEARS THAT LOG_SUM_EXP SHOULD BE STRICTLY CONVEX (equality holds in Cauchy Shwarz only when the x1 and x2 are the same). But STRONG CONVEXITY WHICH REQUIRES A QUADRATIC GAP LOOKS UNLIKELY

Using Midpoint convexity, let us prove that h(x) =log(sum(exp)) is convex (that it is non-increasing is easier)

$$\log\left(\sum_i \exp(x^i)\right)$$

To show

$$\underbrace{h\left(\frac{x_1}{2}+\frac{x_2}{2}\right)}_{\text{LHS}} \le \underbrace{\frac{1}{2}h(x_1) + \frac{1}{2}h(x_2)}_{\text{RHS}}$$

$$\text{RHS} = \frac{1}{2}\log\left(\sum_i \exp(x_1^i)\right) + \frac{1}{2}\log\left(\sum_i \exp(x_2^i)\right)$$

$$= \log\left[\left(\sum_i \underbrace{\exp(x_1^i)}_{\alpha_i^2}\right)^{1/2}\left(\sum_i \underbrace{\exp(x_2^i)}_{\beta_i^2}\right)^{1/2}\right]$$

$$= \log\left[\left(\sum_i \alpha_i^2\right)^{1/2}\left(\sum_i \beta_i^2\right)^{1/2}\right]$$

$$\text{LHS} = \log\left[\sum_i \exp\left(\frac{x_1^i}{2}+\frac{x_2^i}{2}\right)\right]$$

$$= \log\left[\sum_i \left(\exp(x_1^i)\right)^{1/2}\left(\exp(x_2^i)\right)^{1/2}\right]$$

$$= \log \sum_i \alpha_i \beta_i$$

$$\boxed{\log \sum_i \alpha_i\beta_i \le \log\left[\left(\sum_i \alpha_i^2\right)^{1/2}\left(\sum_i \beta_i^2\right)^{1/2}\right]}$$

BY CAUCHY SHWARZ INEQUALITY & THE FACT THAT LOG IS MONOTONICALLY INCREASING FOR POSITIVE ARGUMENTS

Let $\exp\left(\frac{x_1^i}{2}\right) = \alpha_i \qquad \exp\left(\frac{x_2^i}{2}\right) = \beta_i$

$$\boxed{\sum_i \alpha_i\beta_i \le \left[\left(\sum_i \alpha_i^2\right)^{1/2}\left(\sum_i \beta_i^2\right)^{1/2}\right]}$$

CAUCHY SHWARZ INEQUALITY

Which of the Following Loss Functions are Convex?

- L1/L2 Reg Logistic Regression: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \theta^T x_i)) + \lambda \|\theta\|$
- L1/L2 Reg SVMs: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i \theta^T x_i\} + \lambda \|\theta\|$
- L1/L2 Reg Multi-class Logistic Regression:
  $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\} + \sum_{i=1}^{c} \lambda \sum_{j=1}^{m} \|\theta_j\|$
- L1/L2 Reg Least Squares (Lasso): $L(\theta) = \sum_{i=1}^{n} (\theta^T x_i - y_i)^2 + \lambda \|\theta\|$
- Matrix Completion: $L(X) = \sum_{i=1}^{n} ||y_i - A_i(X)||_2^2 + ||X||_*$
- Soft-Max Contextual Bandits: $L(\theta) = \sum_{i=1}^{n} \frac{r_i}{p_i} \frac{\exp(\theta^T x_i^{a_i})}{\sum_{j=1}^{k} \exp(\theta^T x_i^j)} + \lambda \|\theta\|$