Optimization in Machine Learning

Lecture 7: Quasi-convexity, First order and second order convexity conditions

Ganesh Ramakrishnan

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

January, 2025

Recap: Sub-level Sets of Convex Functions

- Lets define *sub-level sets* of a convex function as follows:

Definition

[Sublevel Sets]: Let $\mathcal{D} \subseteq \Re^n$ be a nonempty set and $f : \mathcal{D} \to \Re$. The set

$$L_\alpha(f) = \{\mathbf{x} | \mathbf{x} \in \mathcal{D},\ f(\mathbf{x}) \leq \alpha\}$$

Interiors of Level curves expected to be convex

is called the $\alpha-$sub-level set of f.

Now if a function f is convex,

Recap: Sub-level Sets of Convex Functions

- Lets define *sub-level sets* of a convex function as follows:

Definition

[Sublevel Sets]: Let $\mathcal{D} \subseteq \Re^n$ be a nonempty set and $f : \mathcal{D} \to \Re$. The set

$$L_\alpha(f) = \{\mathbf{x} | \mathbf{x} \in \mathcal{D},\ f(\mathbf{x}) \leq \alpha\}$$

is called the $\alpha-$sub-level set of f.

Now if a function f is convex, its $\alpha-$sub-level set is a convex set.

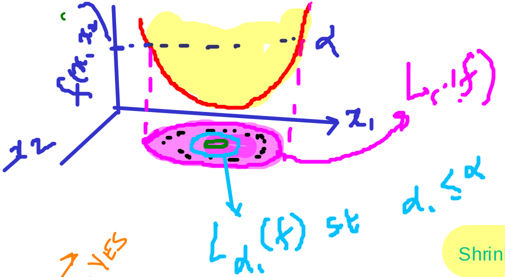
Proof

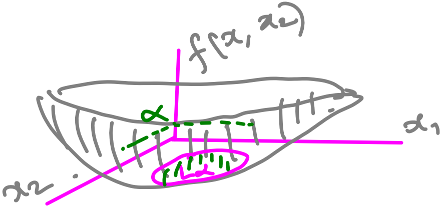
If $x_1, x_2 \in L_\alpha(f)$

$f(\theta x_1 + (1-\theta) x_2) \leq \theta f(x_1) + (1-\theta) f(x_2) \leq \alpha$

$\leq \theta\alpha \qquad \leq (1-\theta)\alpha$

$\Rightarrow \theta x_1 + (1-\theta) x_2 \in L_\alpha(f)$





Shrinking α corresponds to downward movement along the function axis

Ans: YES

Q1: WHAT IF A THE EPIGRAPH IS CONVEX? DOES IT IMPLY THAT THE FUNCTION IS ALSO CONVEX?

Q2: WHAT IF THE SUB-LEVEL SETS ARE CONVEX? DO THEY IMPLY THAT THE FUNCTION IS ALSO CONVEX?

Question: If all (or some) sub-level sets of a function are convex, is it implied that the function itself MUST be convex?

ANS: No.

Proof by counter-example
log |x| for $x > 0$

$$\log|x| \le \alpha_1 \equiv |x| \le e^{\alpha_1}$$

Convex Function $\Rightarrow$ Convex Sub-level sets

Theorem

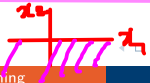
Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \rightarrow \Re$ be a convex function. Then $L_\alpha(f)$ is a convex set for any $\alpha \in \Re$.

Proof: Consider $\mathbf{x}_1, \mathbf{x}_2 \in L_\alpha(f)$. Then by definition of the level set, $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, $f(\mathbf{x}_1) \leq \alpha$ and $f(\mathbf{x}_2) \leq \alpha$. From convexity of $\mathcal{D}$ it follows that for all $\theta \in (0,1)$, $\mathbf{x} = \theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{D}$. Moreover, since f is also convex,

$$f(\mathbf{x}) \leq \theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \leq \theta\alpha + (1-\theta)\alpha = \alpha$$

which implies that $\mathbf{x} \in L_\alpha(f)$. Thus, $L_\alpha(f)$ is a convex set. $\square$

The converse of this theorem does not hold. To illustrate this, consider the function $f(\mathbf{x}) = \frac{x_2}{1+2x_1^2}$. The 0-sublevel set of this function is $\{(x_1, x_2) \mid x_2 \leq 0\}$, which is convex. However, the function $f(\mathbf{x})$ itself is not convex.



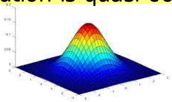
Convex Sub-level sets **DO NOT IMPLY** Convex Function

A function is called quasi-convex if all its sub-level sets are convex sets. Every quasi-convex function is not convex!

Consider the Negative of the normal distribution $-\frac{1}{\sigma\sqrt{2\pi}}exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. This function is quasi-convex but not convex.

: Show that the negative of the normal distribution is quasi-convex

Consider the simpler function $f(x) = -exp(-(x-\mu)^2)$.

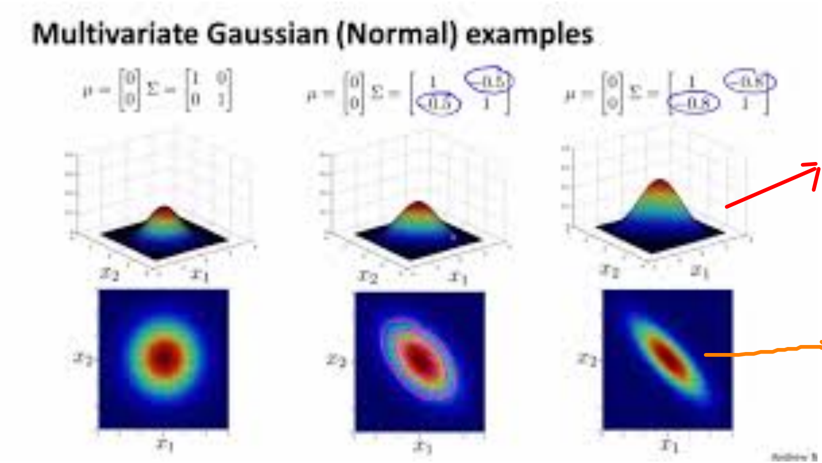


- Then $f'(x) = 2(x-\mu)exp(-(x-\mu)^2)$
- And $f''(x) = 2exp(-(x-\mu)^2) - 4(x-\mu)^2 exp(-(x-\mu)^2) = (2 - 4(x-\mu)^2)exp(-(x-\mu)^2)$ which is < 0 if $(x-\mu)^2 > \frac{1}{2}$,

(Ignoring $\frac{1}{\sqrt{2\pi}}$, set $\sigma^2=1$)

- Thus, the second derivative is negative if $x > \mu + \frac{1}{\sqrt{2}}$ or $x < -\mu - \frac{1}{\sqrt{2}}$.
- Recall from discussion of convexity of $f : \Re \rightarrow \Re$
  - The derivative is not non-decreasing everywhere $\implies$ function is not convex everywhere.

To prove that this function is quasi-convex, we can .... ????????????



Not convex everywhere

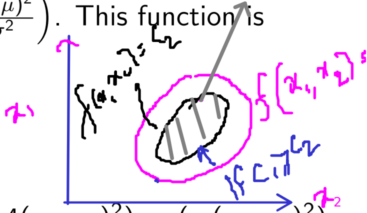
All superlevel sets are convex

Convex Sub-level sets DO NOT IMPLY Convex Function

A function is called quasi-convex if all its sub-level sets are convex sets. Every quasi-convex function is not convex!

Assuming that as c reduced to below c1, the level curves only tend to keep being contained inside, the region inside is called a c1–sub-level set

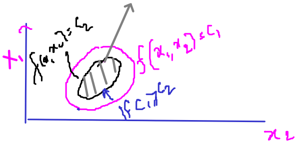
Consider the Negative of the normal distribution $-\frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. This function is quasi-convex but not convex.



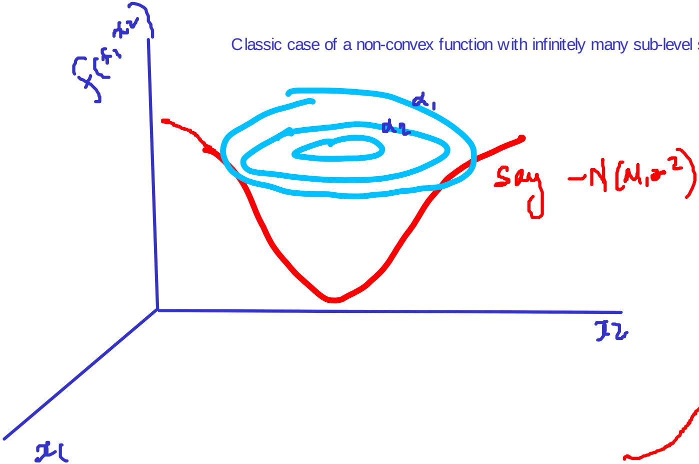
Consider the simpler function $f(x) = -exp(-(x-\mu)^2)$.

- Then $f'(x) = 2(x-\mu)exp(-(x-\mu)^2)$
- And $f''(x) = 2exp(-(x-\mu)^2) - 4(x-\mu)^2 exp(-(x-\mu)^2) = (2-4(x-\mu)^2)exp(-(x-\mu)^2)$ which is < 0 if $(x-\mu)^2 > \frac{1}{2}$,
- Thus, the second derivative is negative if $x > \mu + \frac{1}{\sqrt{2}}$ or $x < -\mu - \frac{1}{\sqrt{2}}$.
- Recall from discussion of convexity of $f : \Re \rightarrow \Re$
  - The derivative is not non-decreasing everywhere $\implies$ function is not convex everywhere.

To prove that this function is quasi-convex, we can ....



Classic case of a non-convex function with infinitely many sub-level sets that are convex



Question: What could be good algorithmic principles for optimizing quasi-convex functions?

1) https://proceedings.neurips.cc/paper/2015/hash/934815ad542a4a7c5e8a2dfa04fea9f5-Abstract.html

Beyond Convexity: Stochastic Quasi-Convex Optimization

Elad Hazan
Princeton University
ehazan@cs.princeton.edu

Kfir Y. Levy
Technion
kfiryl@tx.technion.ac.il

Shai Shalev-Shwartz
The Hebrew University
shais@cs.huji.ac.il

Abstract

Stochastic convex optimization is a basic and well studied primitive in machine learning. It is well known that convex and Lipschitz functions can be minimized efficiently using Stochastic Gradient Descent (SGD).

The Normalized Gradient Descent (NGD) algorithm, is an adaptation of Gradient Descent, which updates according to the direction of the gradients, rather than the gradients themselves. In this paper we analyze a stochastic version of NGD and prove its convergence to a *global* minimum for a wider class of functions: we require the functions to be *quasi-convex* and *locally-Lipschitz*. Quasi-convexity broadens the concept of unimodality to multidimensions and allows for certain types of saddle points, which are a known hurdle for first-order optimization methods such as gradient descent. Locally-Lipschitz functions are only required to be Lipschitz in a small region around the optimum. This assumption circumvents gradient explosion, which is another known hurdle for gradient descent variants.

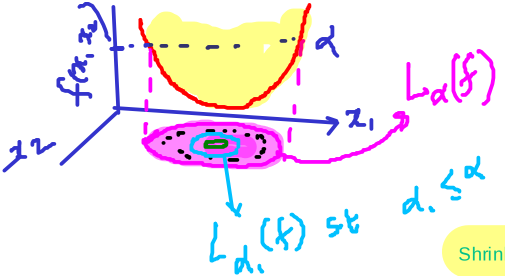
Interestingly, unlike the vanilla SGD algorithm, the stochastic normalized gradient descent algorithm provably requires a minimal minibatch size.

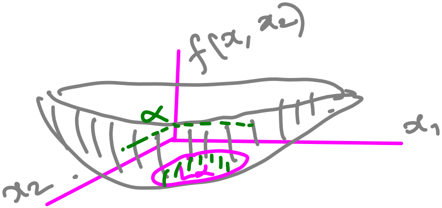
2) SPIDER: Near-Optimal Non-Convex Optimization via Stochastic Path-Integrated Differential Estimator
https://proceedings.neurips.cc/paper/2018/hash/1543843a4723ed2ab08e18053ae6dc5b-Abstract.html
3) AutoML-Zero: Evolving Machine Learning Algorithms From Scratch
http://proceedings.mlr.press/v119/real20a.html
4) Stochastic Variance Reduction for Nonconvex Optimization
https://proceedings.mlr.press/v48/reddi16.html
5) Adaptive Methods for Nonconvex Optimization
https://proceedings.neurips.cc/paper/2018/hash/90365351ccc7437a1309dc64e4db32a3-Abstract.html

Proof that the function is Quasi-Convex

1. Inspect the $L_\alpha(f)$ sublevel sets of this function:
   $L_\alpha(f) = \{x \mid \| - exp(-(x-\mu)^2) \leq \alpha\} = \{x \mid \|exp(-(x-\mu)^2) \geq -\alpha\}$.
2. Since $exp(-(x-\mu)^2)$ is monotonically increasing for $x < \mu$ and monotonically decreasing for $x > \mu$, the set $\{x|exp(-(x-\mu)^2) \geq -\alpha\}$ will be a contiguous closed interval around μ and therefore a convex set.
3. Thus, $f(x) = -exp(-(x-\mu)^2)$ is quasi-convex (and so is its generalization - the negative of the normal density function).

- One can similarly prove that the negative of the multivariate normal density function is also quasi-convex, by inspecting its sub-level sets, which are nothing but **ellipsoids**.





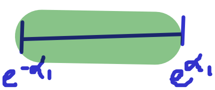
Shrinking α corresponds to downward movement along the function axis

Question: If all (or some) sub-level sets of a function are convex, is it implied that the function itself MUST be convex?

ANS: No.

Proof by counter-example
log |x|

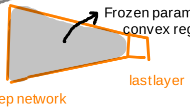
$$\log|x| \le \alpha_1 \equiv |x| \le e^{\alpha_1}$$



Outline

- Understanding the Convexity of Machine Learning Loss Functions [Done]
- Homework Exercises Discussion [Done]
- Basic Subgradient Calculus: Subgradients for non-differentiable convex functions
    - Direction Vector, Directional derivative [Done]
    - Quasi convexity & Sub-level sets of convex functions [Done]
    - Convex Functions & their Epigraphs
    - First-Order Convexity Conditions
    - Subgradients, Subgradient Calculus and Convexity
- Convex Optimization Problems

Outline



Frozen parameters (thatis,with all the pre-final layer parameters frozen,we are often dealing with convex regions ofnon-convex loss functions associated with the deep network

- Understanding the Convexity of Machine Learning Loss Functions [Done]
- Homework Exercises Discussion
- Basic Subgradient Calculus: Subgradients for non-differentiable convex functions
  - Direction Vector, Directional derivative
  - Quasi convexity & Sub-level sets of convex functions
  - Convex Functions & their Epigraphs
  - First-Order Convexity Conditions
  - Subgradients, Subgradient Calculus and Convexity

Applied extensively in convex regions ofnon-convex functions such as in deep learning

- Convex Optimization Problems

Convex Functions and Their Epigraphs

Let us further the connection between convex functions and sets by introducing the concept of the *epigraph* of a function.

Definition

[Epigraph]: Let $\mathcal{D} \subseteq \Re^n$ be a nonempty set and $f : \mathcal{D} \rightarrow \Re$. The set $\{(\mathbf{x}, f(\mathbf{x})|\mathbf{x} \in \mathcal{D}\}$ is called graph of f and lies in $\Re^{n+1}$. The epigraph of f is a subset of $\Re^{n+1}$ and is defined as

$$epi(f) = \{(\mathbf{x}, \alpha)|f(\mathbf{x}) \leq \alpha,\ \mathbf{x} \in \mathcal{D},\ \alpha \in \Re\} \tag{1}$$

In some sense, the epigraph is the set of points lying above the graph of f.

Eg: Recall affine functions of vectors: $\mathbf{a}^T\mathbf{x} + b$ where $\mathbf{a} \in \Re^n$. Its epigraph is $\{(\mathbf{x}, t)|\mathbf{a}^T\mathbf{x} + b \leq t\} \subseteq \Re^{n+1}$ which is a half-space (a convex set).

Convex Functions and Their Epigraphs

Let us further the connection between convex functions and sets by introducing the concept of the *epigraph* of a function.

Definition

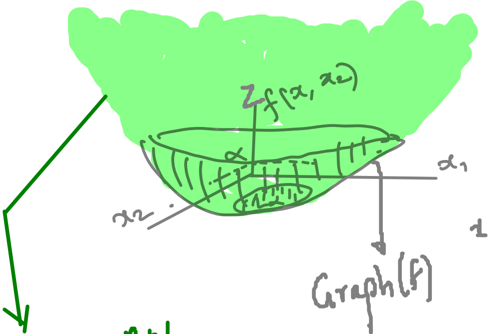
[Epigraph]: Let $\mathcal{D} \subseteq \Re^n$ be a nonempty set and $f : \mathcal{D} \to \Re$. The set $\{(\mathbf{x}, f(\mathbf{x}) | \mathbf{x} \in \mathcal{D}\}$ is called graph of f and lies in $\Re^{n+1}$. The epigraph of f is a subset of $\Re^{n+1}$ and is defined as

$$epi(f) = \{(\mathbf{x}, \alpha) | f(\mathbf{x}) \leq \alpha, \ \mathbf{x} \in \mathcal{D}, \ \alpha \in \Re\} \qquad (1)$$

Homework: Try plotting and inspecting

In some sense, the epigraph is the set of points lying above the graph of f.

Eg: Recall affine functions of vectors: $\mathbf{a}^T\mathbf{x} + b$ where $\mathbf{a} \in \Re^n$. Its epigraph is $\{(\mathbf{x}, t) | \mathbf{a}^T\mathbf{x} + b \leq t\} \subseteq \Re^{n+1}$ which is a half-space (a convex set).



$Epi(f) \subseteq \mathbb{R}^{n+1}$

$(x_1, x_2, \ldots x_n, z)$ is a point in Epi(f)

f is convex IF and ONLY IF its Epi(f) is convex

$L_\alpha(f) \subseteq \mathbb{R}^n$

If $\forall \alpha \in \mathbb{R}$, $L_\alpha(f)$ remains convex
then f is called quasi-convex

Every convex function is quasi-convex (we proved)

Convex Functions and Their Epigraphs (contd)

There is a one to one correspondence between the convexity of function f and that of the set $epi(f)$, as stated in the following result.

Theorem

Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \to \Re$. Then

Convex Functions and Their Epigraphs (contd)

There is a one to one correspondence between the convexity of function f and that of the set $epi(f)$, as stated in the following result.

Theorem

Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \to \Re$. Then f is convex if and only if $epi(f)$ is a convex set.

Proof: f **convex function** $\implies$ $epi(f)$ **convex set**

Proof is very similar to that for sub-level sets

$(x_1, x_2, \ldots x_n, \alpha)$ is a point in Epi(f)

$(x_1', x_2', \ldots x_n', \alpha')$ is a point in Epi(f)

Then $\theta(x_1, x_2 \ldots x_n, \alpha) + (1-\theta)(x_1', x_2' \ldots x_n', \alpha')$

is also a point in epi(f)

Convex Functions and Their Epigraphs (contd)

There is a one to one correspondence between the convexity of function f and that of the set $epi(f)$, as stated in the following result.

Theorem

Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \to \Re$. Then f is convex if and only if $epi(f)$ is a convex set.

Proof: f **convex function** $\implies$ $epi(f)$ **convex set**

Let f be convex. For any $(\mathbf{x}_1, \alpha_1) \in epi(f)$ and $(\mathbf{x}_2, \alpha_2) \in epi(f)$ and any $\theta \in (0,1)$,

$$f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq \theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2)) \leq \theta\alpha_1 + (1-\theta)\alpha_2$$

Since $\mathcal{D}$ is convex, $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{D}$. Therefore, $(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2, \theta\alpha_1 + (1-\theta)\alpha_2) \in epi(f)$. Thus, $epi(f)$ is convex if f is convex. This proves the necessity part.

***epi*(f) convex set $\implies$ f convex function**

To prove sufficiency, assume that $epi(f)$ is convex. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$. So, $(\mathbf{x}_1, f(\mathbf{x}_1)) \in epi(f)$ and $(\mathbf{x}_2, f(\mathbf{x}_2)) \in epi(f)$. Since $epi(f)$ is convex, for $\theta \in (0,1)$,

$$(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2, \theta\alpha_1 + (1-\theta)\alpha_2) \in epi(f)$$

which implies that $f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq \theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2))$ for any $\theta \in (0,1)$. This proves the sufficiency. $\square$

First-Order Convexity Conditions: The complete statement

Theorem

1. *For differentiable* $f : \mathcal{D} \to \Re$ *and convex set* $\mathcal{D}$, f *is convex* **iff**, *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

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

Existence of supporting hypeplane at every point x in terms of the gradient at x

2. f *is* *strictly convex* **iff**, *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, *with* $\mathbf{x} \neq \mathbf{y}$,

$$f(\mathbf{y}) \boxed{>} f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

3. f *is* *strongly convex* **iff**, *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, *and for some constant* $c > 0$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \boxed{\frac{1}{2}c||\mathbf{y} - \mathbf{x}||^2}$$

First-Order Convexity Conditions: The complete statement

Theorem

1. *For differentiable $f : \mathcal{D} \rightarrow \Re$ and convex set $\mathcal{D}$, f is convex* **iff**, *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

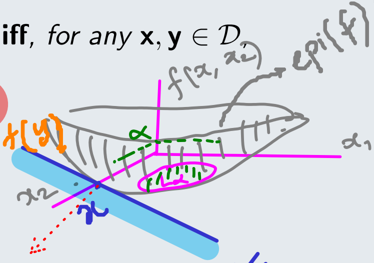
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

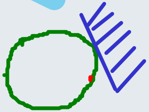
2. *f is strictly convex* **iff**, *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, *with* $\mathbf{x} \neq \mathbf{y}$,

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

3. *f is strongly convex* **iff**, *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, *and for some constant* $c > 0$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}c||\mathbf{y} - \mathbf{x}||^2$$





Does this non-convex set have a supporting hyperplane at x ?

First-Order Convexity Condition: Proof

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (1). Suppose (1) holds. Consider $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ and any $\theta \in (0, 1)$. Let $\mathbf{x} = \theta \mathbf{x}_1 + (1 - \theta)\mathbf{x}_2$. Then, $f(\mathbf{x}_1) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x})$ and $f(\mathbf{x}_2) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x})$

First-Order Convexity Condition: Proof

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (1). Suppose (1) holds. Consider $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ and any $\theta \in (0,1)$. Let $\mathbf{x} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$. Then, $f(\mathbf{x}_1) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x})$ and $f(\mathbf{x}_2) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x})$
Adding $(1-\theta)$ times the second inequality to θ times the first, we get,

First-Order Convexity Condition: Proof

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (1). Suppose (1) holds. Consider $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ and any $\theta \in (0,1)$. Let $\mathbf{x} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$. Then, $f(\mathbf{x}_1) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x})$ and $f(\mathbf{x}_2) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x})$
Adding $(1-\theta)$ times the second inequality to θ times the first, we get,

$$\theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \geq f(\mathbf{x})$$

which proves that $f(\mathbf{x})$ is a convex function. In the case of strict convexity,

First-Order Convexity Condition: Proof

Sufficiency: The proof of sufficiency is very similar for all the three statements of the theorem. So we will prove only for statement (1). Suppose (1) holds. Consider $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ and any $\theta \in (0,1)$. Let $\mathbf{x} = \theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2$. Then, $f(\mathbf{x}_1) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_1 - \mathbf{x})$ and $f(\mathbf{x}_2) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{x}_2 - \mathbf{x})$
Adding $(1-\theta)$ times the second inequality to θ times the first, we get,

$$\theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \geq f(\mathbf{x})$$

which proves that $f(\mathbf{x})$ is a convex function. In the case of strict convexity, strict inequality holds in (2) and it follows through. In the case of strong convexity, we obtain (after some manipulation): $\theta[f(x_1) - c/2\|x_1\|^2] + (1-\theta)[f(x_2) - c/2\|x_2\|^2] \geq f(x) - c/2\|x\|^2$ which implies that $f(x) - c/2\|x\|^2$ is convex!

First-Order Convexity Conditions: Proofs

Necessity: Suppose f is convex. Then for all $\theta \in (0,1)$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, we must have

$$f(\theta \mathbf{x}_2 + (1-\theta)\mathbf{x}_1) \leq \theta f(\mathbf{x}_2) + (1-\theta) f(\mathbf{x}_1)$$

Thus,

$$\nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) =$$

First-Order Convexity Conditions: Proofs

Necessity: Suppose f is convex. Then for all $\theta \in (0,1)$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, we must have

$$f(\theta \mathbf{x}_2 + (1-\theta)\mathbf{x}_1) \leq \theta f(\mathbf{x}_2) + (1-\theta) f(\mathbf{x}_1)$$

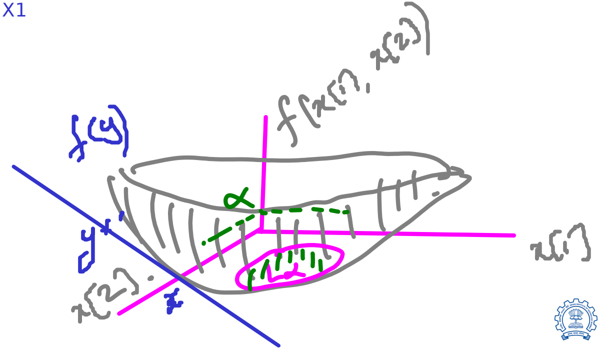
Thus,

$\nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) =$ THE SCALED VERSION OF DIRECTIONAL DERIVATIVE OF f AT X1 ALONG X2 - X1

THIS HYPERPLANE REPRESENTS THE MAGNITUDE OF

$$\lim_{\theta \to 0} \frac{f(x+\theta(y-x)) - f(x)}{\theta}.$$

ALONG y-x



First-Order Convexity Conditions: Proofs

Necessity: Suppose f is convex. Then for all $\theta \in (0,1)$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, we must have

$$f(\theta \mathbf{x}_2 + (1-\theta)\mathbf{x}_1) \leq \theta f(\mathbf{x}_2) + (1-\theta) f(\mathbf{x}_1)$$

Thus,

$$\nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) = \lim_{\theta \to 0} \frac{f\left(\mathbf{x}_1 + \theta(\mathbf{x}_2 - \mathbf{x}_1)\right) - f\left(\mathbf{x}_1\right)}{\theta}$$

Necessity: Suppose f is convex. Then for all $\theta \in (0,1)$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, we must have

$$f(\theta \mathbf{x}_2 + (1-\theta)\mathbf{x}_1) \leq \theta f(\mathbf{x}_2) + (1-\theta) f(\mathbf{x}_1)$$

Thus,

$$\nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) = \lim_{\theta \to 0} \frac{f\left(\mathbf{x}_1 + \theta(\mathbf{x}_2 - \mathbf{x}_1)\right) - f\left(\mathbf{x}_1\right)}{\theta} \leq f(\mathbf{x}_2) - f(\mathbf{x}_1)$$

This proves necessity for (1). The necessity proofs for (2) and (3) are very similar, except for a small difference for the case of strict convexity; the strict inequality is not preserved when we take limits. Suppose equality does hold in the case of strict convexity, that is for a strictly convex function f, let

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1)$$

for some $\mathbf{x}_2 \neq \mathbf{x}_1$.

First-Order Convexity Conditions: Proofs

Necessity (contd for strict case):
Because f is strictly convex, for any $\theta \in (0,1)$ we can write

$$f(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) = f(\mathbf{x}_2 + \theta(\mathbf{x}_1 - \mathbf{x}_2)) < \theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2)$$

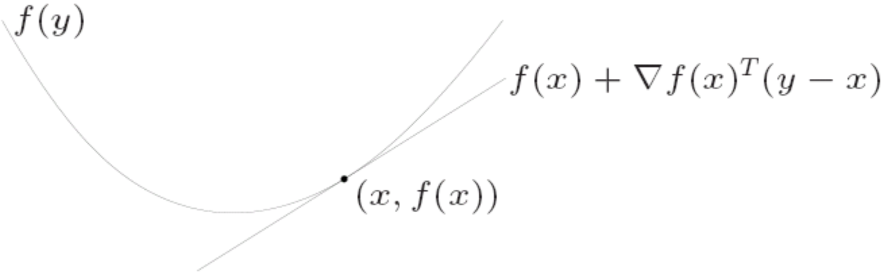
Since (1) is already proved for convex functions, we use it in conjunction with the previous two expressions to obtain

$$f(\mathbf{x}_2) + \theta \nabla^T f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2) \le f(\mathbf{x}_2 + \theta(\mathbf{x}_1 - \mathbf{x}_2)) < f(\mathbf{x}_2) + \theta \nabla^T f(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)$$

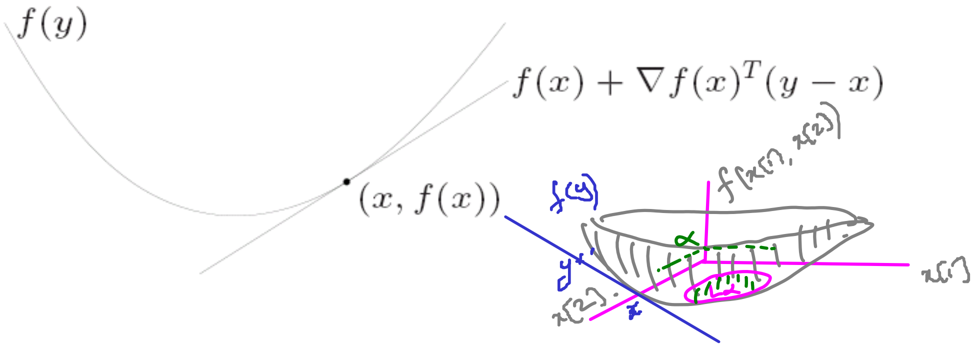
which is a contradiction. Thus, equality can never hold in (2) for any $\mathbf{x}_1 \neq \mathbf{x}_2$. This proves the necessity of (2). (3) can be proved by using the fact that $g(x) = f(x) - c/2\|x\|$ is convex and then applying (1) to g.

First-Order Convexity Conditions: The complete statement

The geometrical interpretation of this theorem is that at any point, the linear approximation based on a local derivative gives a lower estimate of the function, *i.e.* the convex function always lies above the supporting hyperplane at that point. This is pictorially depicted below:



The geometrical interpretation of this theorem is that at any point, the linear approximation based on a local derivative gives a lower estimate of the function, *i.e.* the convex function always lies above the supporting hyperplane at that point. This is pictorially depicted below:



Second Order Conditions of Convexity

- Recall the Hessian of a continuous function:

$$\nabla^2 f(w) = \begin{pmatrix} \frac{\partial^2 f}{\partial w_1^2} & \frac{\partial^2 f}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 f}{\partial w_1 \partial w_n} \\ \frac{\partial^2 f}{\partial w_2 \partial w_1} & \frac{\partial^2 f}{\partial w_2^2} & \cdots & \frac{\partial^2 f}{\partial w_2 \partial w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial w_n \partial w_1} & \frac{\partial^2 f}{\partial w_n \partial w_2} & \cdots & \frac{\partial^2 f}{\partial w_n^2} \end{pmatrix}$$

- f is convex if and only if, a) $dom(f)$ is convex, and for all $x \in dom(f)$, $\nabla^2 f(x) \geq 0$ (i.e. $\nabla^2 f(x)$ is positive semi-definite).

Second Order Conditions of Convexity

Can we use the Hessian to prove that the logSumExp function is Convex?
Answer is YES
Boyd's book uses the fact that Hessian being positive semi-definite is necessary and sufficient condition for convexity

- Recall the Hessian of a continuous function:

$$\frac{\partial^2 \exp(x)}{\partial x^2} = \exp(x)$$

$$\text{sumexp} = \sum_i \exp(x_i)$$

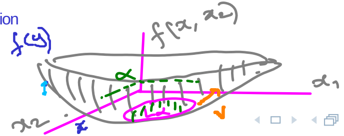
$$\nabla^2 \text{sumexp} = \begin{bmatrix} \exp(x_i) & & 0 \\ & \ddots & \\ 0 & & \ddots \end{bmatrix}$$

$$\nabla^2 f(w) = \begin{pmatrix} \frac{\partial^2 f}{\partial w_1^2} & \frac{\partial^2 f}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 f}{\partial w_1 \partial w_n} \\ \frac{\partial^2 f}{\partial w_2 \partial w_1} & \frac{\partial^2 f}{\partial w_2^2} & \cdots & \frac{\partial^2 f}{\partial w_2 \partial w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial w_n \partial w_1} & \frac{\partial^2 f}{\partial w_n \partial w_2} & \cdots & \frac{\partial^2 f}{\partial w_n^2} \end{pmatrix}$$

INTUITION:
1) First order condition: The directional derivative is non-decreasing in every direction
2) Second order condition: The curvature is positive in every direction!

- f is convex if and only if, a) $dom(f)$ is convex, and for all $x \in dom(f)$, $\nabla^2 f(x) \geq 0$ (i.e. $\nabla^2 f(x)$ is positive semi-definite).

To show that LogSumExp is convex, can we prove that the quadrative expression is always non-negative



$\forall v.\ \boxed{v^T \nabla^2 f(x) v \geq 0}$

$v^T \nabla^2 \text{logSumExp}\ V =$ EXPAND AS HOMEWORK!