CS6301: Optimization in Machine Learning

Lecture 3 & 4: Convexity and Convex Optimization

Rishabh Iyer

Department of Computer Science
University of Texas, Dallas
https://sites.google.com/view/cs-6301-optml/home

January 22, 2020

Outline

- Recap from Previous Lecture
- Basics of Convexity: Convex Sets and Convex Functions
- Properties and Examples of Convex functions
- Basic Subgradient Calculus: Subgradients for non-differentiable convex functions
- Understanding the Convexity of Machine Learning Loss Functions
- Convex Optimization Problems

Recap From Previous Lecture

- Review of Notation: Vectors and Matrices
- Derivatives, Partial Derivatives, Gradients and Hessians
- Implementing Loss Functions and Gradients in Python.
- Assignment 1 was posted last week. The due date for this assignment is January 31st. This can be a time consuming assignment so please start early.
- Feel free to ask me any questions after class or during my office hours.
- **Hopefully all of you have started the assignments and are halfway through!**

Recap: Logistic Regression Gradient

- Lets start with Regularized Logistic Regression. Assume the Labels $y_i \in \{-1, +1\}$.
- The objective of Reg Logistic Loss is:

$$L(w) = \lambda/2\|w\|^2 + \sum_{i=1}^{n} \log(1 + \exp(-y_i w^T x_i)) \tag{1}$$

- Compute the gradient of this Loss?
- Gradient:

$$\begin{aligned}\nabla L(w) =& \lambda w + \sum_{i=1}^{n} \frac{-y_i \exp(-y_i(w^T x_i))}{1 + \exp(-y_i w^T x_i)} x_i \\ =& \lambda w + \sum_{i=1}^{n} \frac{-y_i}{1 + \exp(y_i w^T x_i)} x_i\end{aligned}$$

Recap: Logistic Regression Hessian

- Lets next compute the Hessian.
- Recall the Gradient:

$$\nabla L(w) = \lambda w + \sum_{i=1}^{n} \frac{-y_i}{1 + \exp(y_i w^T x_i)} x_i$$

- You can derive the Hessian as:

$$\nabla^2 L(w) = \lambda I + \sum_{i=1}^{n} \frac{\exp(y_i w^T x_i)}{(1 + \exp(y_i w^T x_i))^2} x_i x_i^T$$

- Define $\sigma(z) = 1/(1 + \exp(-z))$. Then its easy to see that:

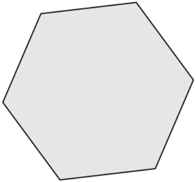
$$\nabla^2 L(w) = \sigma(y_i w^T x_i)(1 - \sigma(y_i w^T x_i)) x_i x_i^T + \lambda I$$

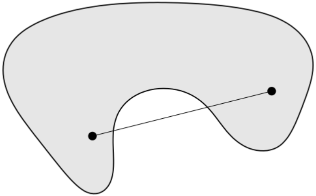
Numerical Issues and Implementations

- We studied numerical issues with $\log(1 + \exp(-x))$
- How do we fix it? See question in Assignment 1!
- Also numerical issues with $\log(\exp(x1) + \exp(x2))$ or $\exp(x1)/(\exp(x1) + \exp(x2))$
- We also covered how to implement the Logistic Loss Function.

Convex Sets

A set C is a **convex set** if the line segment between any two points of C lies in C, i.e. if for any $x, y \in C$ and for any $0 < \lambda < 1$, we have that $\lambda x + (1 - \lambda) y \in C$.





Source: Boyd's Textbook

Properties of Convex Sets

- Intersections of Convex Sets are Convex. Let $C_1, \cdots, C_k$ be convex sets, then $\cap_{i=1}^k C_i$ is convex.
- Is the union of convex sets convex?
- Projections onto convex sets are unique (and often efficient to compute).

$$P_C(x) = \text{argmin}_{y \in C} \|y - x\|$$

- Examples of Convex Sets:
  - $C = \{x \in \mathbb{R}^n : \|x\| \leq k\}$
  - $C = \{x \in \mathbb{R}^n : w^T x \leq k\}$
  - Given a convex function f, the associated set $C_f = \{x \in \mathbb{R}^n : f(x) \leq k\}$ is convex.

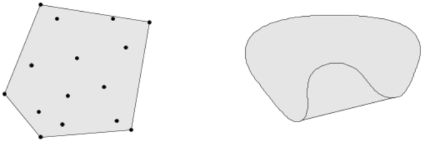
Convex combination and convex hull

- **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.



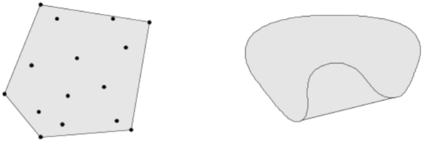
- Should S be always convex?
- What about the convexity of conv(S)?

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

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

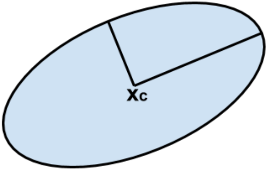
- **Convex hull or conv(S)** is the set of all convex combinations of point in the set S.



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

Euclidean balls and ellipsoids

- **Euclidean ball** with **center** $\mathbf{x}_c$ and **radius** r is given by:
  $B(\mathbf{x}_c, r) = \{\mathbf{x} — \|\mathbf{x} - \mathbf{x}_c\|_2 \leq r\} = \{\mathbf{x}_c + ru — \|u\|_2 \leq 1 \}$
- **Ellipsoid** is a **set** of form:
  $\{\mathbf{x} — (\mathbf{x} - \mathbf{x}_c)^T P^{-1}(\mathbf{x} - \mathbf{x}_c) \leq 1 \}$, where P $\in S^n_{++}$ i.e. P is SPD matrix.
  - Other representation: $\{\mathbf{x}_c$ + A u — $\|u\|_2 \leq 1\}$ with A square and non-singular(i.e. A^{-1} exists).



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: $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}_x\| \leq r\}$ is a convex set. Why?

[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: $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}_x\| \leq r\}$ 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.)

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.

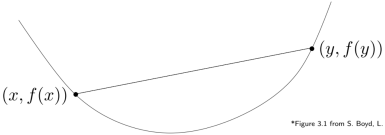
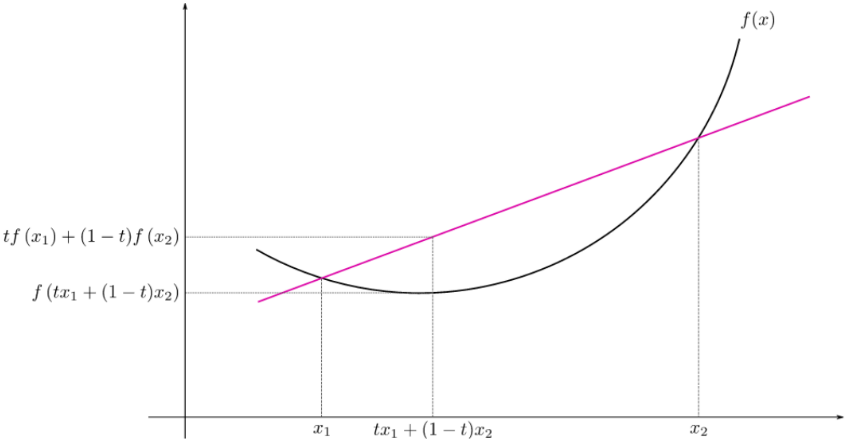


*Figure 3.1 from S. Boyd, L. Vandenberghe

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

Intuition of Convexity



Equivalent Definitions of Convex Functions

The following conditions are equivalent (in one dimension) when $dom(f)$ is a convex set:

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

Equivalent Definitions of Convex Functions

The following conditions are equivalent (in one dimension) when $dom(f)$ is a convex set:

1. f is convex iff for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$
2. f is convex iff $\forall x_1, x_2, x_3$ such that $x_1 < x_2 < x_3$ it holds that $\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_2)}{x_3-x_2}$

Equivalent Definitions of Convex Functions

The following conditions are equivalent (in one dimension) when $dom(f)$ is a convex set:

1. f is convex iff for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
2. f is convex iff $\forall x_1, x_2, x_3$ such that $x_1 < x_2 < x_3$ it holds that $\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_2)}{x_3-x_2}$
3. f is convex iff $f'(x)$ is a monotonic function of x. In other words, $f'(x_2) \geq f'(x_1)$ if $x_2 \geq x_1$.

Equivalent Definitions of Convex Functions

The following conditions are equivalent (in one dimension) when $dom(f)$ is a convex set:

1. f is convex iff for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
2. f is convex iff $\forall x_1, x_2, x_3$ such that $x_1 < x_2 < x_3$ it holds that $\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_2)}{x_3-x_2}$
3. f is convex iff $f'(x)$ is a monotonic function of x. In other words, $f'(x_2) \geq f'(x_1)$ if $x_2 \geq x_1$.
4. f is convex iff $f''(x) \geq 0$

Are the following functions convex?

- $f(x) = \exp(x)$
- $f(x) = \exp(-x)$
- $f(x) = \log x$
- $f(x) = \sin x$
- $f(x) = \log(1 + \exp(-x))$
- $f(x) = x^2$
- $f(x) = x^{2n}$ where n is an integer
- $f(x) = \max\{x, 0\}$
- $f(x) = \sqrt{x}$

From 1 dimensions to n dimensions

- Conditions for convexity in 1 dimensions is eas(ier)
- In the rest of this lecture, we shall understand how to extend this to n dimensions.
- Note that the basic definition of convexity still holds: f is convex iff for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$
- We shall look at some results which will help us prove some functions are convex!

Strongly Convex Functions

- 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
- 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!
- To summarize: Strong Convexity $\Rightarrow$ Strict Convexity $\Rightarrow$ Convexity! (The converse does not hold)

Convex Functions

Examples of Convex Functions

- Linear Functions: $f(x) = a^T x$

Convex Functions

Examples of Convex Functions

- Linear Functions: $f(x) = a^T x$
- Affine Functions: $f(x) = a^T x + b$

Convex Functions

Examples of Convex Functions

- Linear Functions: $f(x) = a^T x$
- Affine Functions: $f(x) = a^T x + b$
- Exponential: $f(x) = exp(\alpha x)$

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

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

- **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 $\alpha_i \geq 0, 1 \leq i \leq n$.

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:

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.

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 $h : \mathbb{R} \to \mathbb{R}$.

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

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

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: Take double derivative and try to show that $\nabla^2 f \geq 0$ (easier to prove this for $m = 1$).

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: Take double derivative and try to show that $\nabla^2 f \geq 0$ (easier to prove this for $m = 1$).
- Examples:

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: Take double derivative and try to show that $\nabla^2 f \geq 0$ (easier to prove this for $m = 1$).
- Examples:
  - $f(x) = \exp(f(x))$ is convex if f 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: Take double derivative and try to show that $\nabla^2 f \geq 0$ (easier to prove this for $m = 1$).
- Examples:
  - $f(x) = \exp(f(x))$ is convex if f is convex
  - $1/g(x)$ is convex if g is concave.

Composition with Vector Functions

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

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

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

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

- 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

Composition with Vector Functions

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

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

- 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:

Composition with Vector Functions

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

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

- 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:
  - $f(x) = \sum_i \log(g(x))$ is concave if g is concave and positive

Composition with Vector Functions

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

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

- 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:
  - $f(x) = \sum_i \log(g(x))$ is concave if g is concave and positive
  - $\log \sum_{i=1}^k \exp(g_i(x))$ is convex if g_i is convex.

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\left\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\right\}$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is convex. For example:
  - The function that returns the maximum eigenvalue of a symmetric matrix X, *viz.*, $\lambda_{max}(X) = \sup_{\mathbf{y} \in \mathcal{S}} \frac{\|X\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$ is

UT DALLAS

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\left\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\right\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup\limits_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is convex. For example:
  - The function that returns the maximum eigenvalue of a symmetric matrix X, *viz.*, $\lambda_{max}(X) = \sup\limits_{\mathbf{y} \in \mathcal{S}} \frac{\|X\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup\limits_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is convex. For example:
  - The function that returns the maximum eigenvalue of a symmetric matrix X, *viz.*, $\lambda_{max}(X) = \sup\limits_{\mathbf{y} \in \mathcal{S}} \frac{\|X\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$ is a convex function of the symmetrix matrix X.

Which of the Following Loss Functions are Convex?

- L1/L2 Reg Logistic Regression:
  $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \theta^T x_i)) + \lambda \|\theta\|$

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: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i \theta^T x_i\} + \lambda \|\theta\|$

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: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\} + \sum_{i=1}^{c} \lambda \sum_{j=1}^{m} \|\theta_j\|$

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): $L(\theta) = \sum_{i=1}^n (\theta^T x_i - y_i)^2 + \lambda\|\theta\|$

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: $L(X) = \sum_{i=1}^{n} ||y_i - A_i(X)||_2^2 + ||X||_*$

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: $L(\theta) = \sum_{i=1}^{n} \frac{r_i}{p_i} \frac{\exp(\theta^T x_i^{a_i})}{\sum_{j=1}^{k} \exp(\theta^T x_i^j)} + \lambda\|\theta\|$

The Direction Vector

- Consider a function $f(\mathbf{x})$, with $\mathbf{x} \in \Re^n$.
- We start with the concept of the direction at a point $\mathbf{x} \in \Re^n$.
- We will represent a vector by $\mathbf{x}$ and the k^{th} component of $\mathbf{x}$ by x_k.
- Let $\mathbf{u}^k$ be a unit vector pointing along the k^{th} coordinate axis in $\Re^n$;
- $u_k^k = 1$ and $u_j^k = 0, \ \forall j \neq k$
- An arbitrary direction vector $\mathbf{v}$ at $\mathbf{x}$ is a vector in $\Re^n$ with unit norm (*i.e.*, $||\mathbf{v}|| = 1$) and component v_k in the direction of $\mathbf{u}^k$.

Directional derivative and the gradient vector

Let $f : \mathcal{D} \to \Re$, $\mathcal{D} \subseteq \Re^n$ be a function.

Definition

[Directional derivative]: The *directional derivative* of $f(\mathbf{x})$ at $\mathbf{x}$ in the direction of the unit vector $\mathbf{v}$ is

Directional derivative and the gradient vector

Let $f : \mathcal{D} \rightarrow \Re$, $\mathcal{D} \subseteq \Re^n$ be a function.

Definition

[Directional derivative]: The *directional derivative* of $f(\mathbf{x})$ at $\mathbf{x}$ in the direction of the unit vector $\mathbf{v}$ is

$$D_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} \tag{2}$$

provided the limit exists.

Directional Derivative

As a special case, when $\mathbf{v} = \mathbf{u}^k$ the directional derivative reduces to the partial derivative of f with respect to x_k.

$$D_{\mathbf{u}^k} f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_k}$$

If $f(\mathbf{x})$ is a differentiable function of $\mathbf{x} \in \Re^n$, then f has a directional derivative in the direction of any unit vector $\mathbf{v}$, and

$$D_{\mathbf{v}} f(\mathbf{x}) = \sum_{k=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_k} v_k = \nabla f^T v \tag{3}$$

Sublevel 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,

Sublevel 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.

Convex Function $\Rightarrow$ Convex Sub-level sets

Theorem

Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \to \Re$ 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 Function $\Rightarrow$ Convex Sub-level sets

Theorem

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

A function is called quasi-convex if all its sub-level sets are convex

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.

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 $f: \Re \to \Re$ if the derivative is not non-decreasing everywhere $\implies$ function is not convex everywhere.

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

Proof that the function is Quasi-Convex

1. Inspect the $L_\alpha(f)$ sublevel sets of this function: $L_\alpha(f) = \{x| - exp(-(x-\mu)^2) \leq \alpha\} = \{x|exp(-(x-\mu)^2) \geq -\alpha\}$.
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**.

Proof that the function is Quasi-Convex

1. Inspect the $L_\alpha(f)$ sublevel sets of this function: $L_\alpha(f) = \{x| - exp(-(x-\mu)^2) \leq \alpha\} = \{x|exp(-(x-\mu)^2) \geq -\alpha\}$.
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**.

UT DALLAS

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

$$epi(f) = \{(\mathbf{x}, \alpha)|f(\mathbf{x}) \leq \alpha,\ \mathbf{x} \in \mathcal{D},\ \alpha \in \Re\} \tag{4}$$

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 (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 $f : \mathcal{D} \rightarrow \Re$. 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**

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* $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

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

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

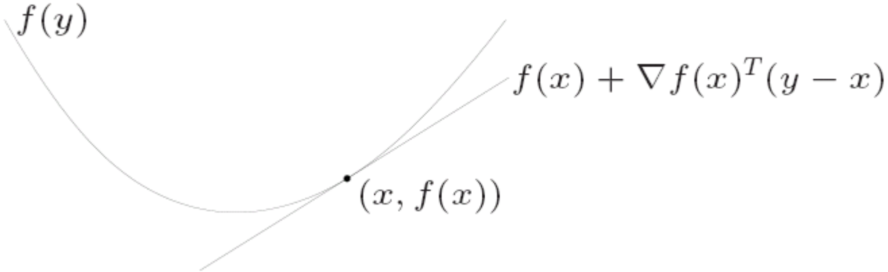
$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

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

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

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:



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,

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

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,

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

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

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) \tag{5}$$

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

$$f(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) = f(\mathbf{x}_2 + \theta(\mathbf{x}_1 - \mathbf{x}_2)) < \theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \quad (6)$$

Since (1) is already proved for convex functions, we use it in conjunction with (5), and (6), to get

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

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.

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 $x \in dom(f)$, $\nabla^2 f(x) \succcurlyeq 0$ (i.e. $\nabla^2 f(x)$ is positive semi-definite).
- In one dimension, this means f is convex iff $f''(x) \geq 0$

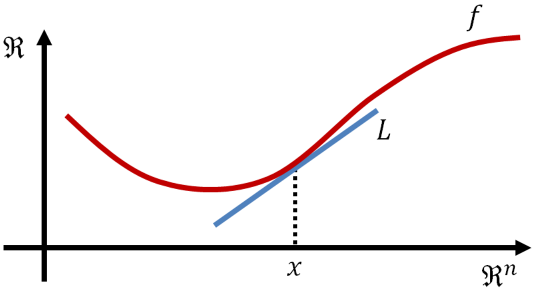
Monotonicity of Gradients

Theorem

A function f is convex if and only if $dom(f)$ is convex and for all $x, y \in dom(f), (\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$

- This directly follows from the first order characterization of convexity
- Note that $f(x) \geq f(y) + \nabla f(y)^T (x - y)$ and $f(y) \geq f(x) + \nabla f(x)^T (y - x)$.
- Adding both the inequalities above we get the result!
- Note that the 1D monotonicity statement we saw earlier in the class is a special case of this!

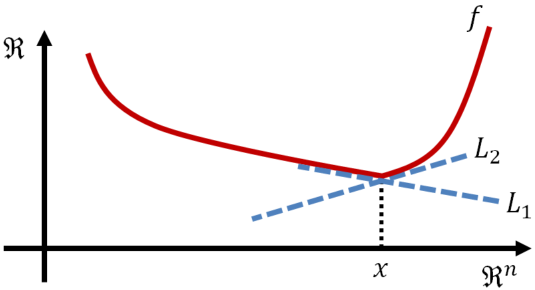
(Sub)Gradients and Convexity (contd)



To say that a function $f : \Re^n \mapsto \Re$ is differentiable at $\mathbf{x}$ is to say that there is a single unique linear tangent that under estimates the function:

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

(Sub)Gradients and Convexity (contd)

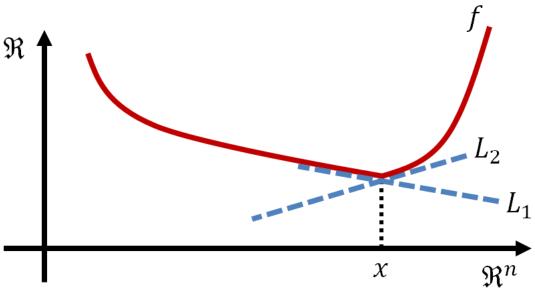


In this figure we see the function f at $\mathbf{x}$ has many possible linear tangents that may fit appropriately. Then a **subgradient** is any $\mathbf{h} \in \Re^n$ (same dimension as x) such that:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{h}^T(\mathbf{y} - \mathbf{x}), \ \forall \mathbf{y}$$

Thus, intuitively, if a function is differentiable at a point $\mathbf{x}$ then

(Sub)Gradients and Convexity (contd)



In this figure we see the function f at $\mathbf{x}$ has many possible linear tangents that may fit appropriately. Then a **subgradient** is any $\mathbf{h} \in \Re^n$ (same dimension as x) such that:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{h}^T(\mathbf{y} - \mathbf{x}), \ \ \forall \mathbf{y}$$

Thus, intuitively, if a function is differentiable at a point $\mathbf{x}$ then it has a unique subgradient at that point ($\bigtriangledown f(\mathbf{x})$). Formal Proof?

Detour: Convexity and Continuity

- Let f be a convex function and suppose $dom(f)$ is open. Then f is continuous.
- How *wild* can non-differentiable convex functions be?
- While there are continuous functions which are nowhere differentiable, (see `https://en.wikipedia.org/wiki/Weierstrass_function`), convex functions cannot be pathological!
- Infact, a convex function is differentiable *almost* everywhere. In other words, the set of points where f is non-differentiable is of measure 0.
- However we cannot ignore the non-differentiability, since a) the global minima could easily be a point of non differentiability and b) with any optimization algorithms, you can stumble upon these "kinks".

(Sub)Gradients and Convexity (contd)

- A **subdifferential** is the closed convex set of all subgradients of the convex function f:

$$\partial f(\mathbf{x}) = \{\mathbf{h} \in \Re^n : \mathbf{h} \text{ is a subgradient of } f \text{ at } \mathbf{x}\}$$

Note that this set is guaranteed to be nonempty unless f is not convex.

(Sub)Gradients and Convexity (contd)

- A **subdifferential** is the closed convex set of all subgradients of the convex function f:

$$\partial f(\mathbf{x}) = \{\mathbf{h} \in \Re^n : \mathbf{h} \text{ is a subgradient of } f \text{ at } \mathbf{x}\}$$

  Note that this set is guaranteed to be nonempty unless f is not convex.

- **Pointwise Maximum:**. if $f(\mathbf{x}) = max_{i=1\ldots m} f_i(\mathbf{x})$, then $\partial f(\mathbf{x}) = conv\left(\bigcup_{i:f_i(\mathbf{x})=f(\mathbf{x})} \partial f_i(\mathbf{x})\right)$, which is the convex hull of union of subdifferentials of all active functions at x.

(Sub)Gradients and Convexity (contd)

- A **subdifferential** is the closed convex set of all subgradients of the convex function f:

$$\partial f(\mathbf{x}) = \{\mathbf{h} \in \Re^n : \mathbf{h} \text{ is a subgradient of } f \text{ at } \mathbf{x}\}$$

  Note that this set is guaranteed to be nonempty unless f is not convex.

- **Pointwise Maximum:**. if $f(\mathbf{x}) = max_{i=1\dots m} f_i(\mathbf{x})$, then $\partial f(\mathbf{x}) = conv\left(\bigcup_{i: f_i(\mathbf{x}) = f(\mathbf{x})} \partial f_i(\mathbf{x})\right)$, which is the convex hull of union of subdifferentials of all active functions at x.

- **General pointwise maximum:** if $f(\mathbf{x}) = max_{s \in S} f_s(\mathbf{x})$, then under some regularity conditions (on S, f_s), $\partial f(\mathbf{x}) = cl\left\{conv\left(\bigcup_{s: f_s(\mathbf{x}) = f(\mathbf{x})} \partial f_s(\mathbf{x})\right)\right\}$

Subgradient of $\|\mathbf{x}\|_1$

Assume $\mathbf{x} \in \Re^n$. Then

- $\|\mathbf{x}\|_1 =$

Subgradient of $\|\mathbf{x}\|_1$

Assume $\mathbf{x} \in \Re^n$. Then

- $\|\mathbf{x}\|_1 = \max_{\mathbf{s} \in \{-1,+1\}^n} \mathbf{x}^T\mathbf{s}$ which is a pointwise maximum of 2^n functions

Subgradient of $\|\mathbf{x}\|_1$

Assume $\mathbf{x} \in \Re^n$. Then

- $\|\mathbf{x}\|_1 = \max_{\mathbf{s} \in \{-1,+1\}^n} \mathbf{x}^T\mathbf{s}$ which is a pointwise maximum of 2^n functions
- Let $\mathcal{S}^* \subseteq \{-1,+1\}^n$ be the set of $\mathbf{s}$ such that for each $\mathbf{s} \in \mathcal{S}^*$, the value of $\mathbf{x}^T\mathbf{s}$ is the same max value.

Subgradient of $\|\mathbf{x}\|_1$

Assume $\mathbf{x} \in \Re^n$. Then

- $\|\mathbf{x}\|_1 = \max_{\mathbf{s} \in \{-1,+1\}^n} \mathbf{x}^T\mathbf{s}$ which is a pointwise maximum of 2^n functions
- Let $\mathcal{S}^* \subseteq \{-1, +1\}^n$ be the set of $\mathbf{s}$ such that for each $\mathbf{s} \in \mathcal{S}^*$, the value of $\mathbf{x}^T\mathbf{s}$ is the same max value.
- Thus, $\partial\|\mathbf{x}\|_1 = conv\left(\bigcup_{\mathbf{s}\in\mathcal{S}^*} \mathbf{s}\right)$.

More of Basic Subgradient Calculus

- Scaling: $\partial(af) = a \cdot \partial f$ provided $a > 0$. The condition $a > 0$ makes function f remain convex.
- Addition: $\partial(f_1 + f_2) = \partial(f_1) + \partial(f_2)$
- Affine composition: if $g(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b})$, then $\partial g(\mathbf{x}) = A^T \partial f(A\mathbf{x} + b)$
- Norms: important special case, $f(\mathbf{x}) = ||\mathbf{x}||_p$

More of Basic Subgradient Calculus

- Scaling: $\partial(af) = a \cdot \partial f$ provided $a > 0$. The condition $a > 0$ makes function f remain convex.
- Addition: $\partial(f_1 + f_2) = \partial(f_1) + \partial(f_2)$
- Affine composition: if $g(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b})$, then $\partial g(\mathbf{x}) = A^T \partial f(A\mathbf{x} + b)$
- Norms: important special case, $f(\mathbf{x}) = ||\mathbf{x}||_p = \max_{||\mathbf{z}||_q \leq 1} \mathbf{z}^T\mathbf{x}$ where q is such that $1/p + 1/q = 1$. Then

More of Basic Subgradient Calculus

- Scaling: $\partial(af) = a \cdot \partial f$ provided $a > 0$. The condition $a > 0$ makes function f remain convex.
- Addition: $\partial(f_1 + f_2) = \partial(f_1) + \partial(f_2)$
- Affine composition: if $g(\mathbf{x}) = f(A\mathbf{x} + \mathbf{b})$, then $\partial g(\mathbf{x}) = A^T \partial f(A\mathbf{x} + b)$
- Norms: important special case, $f(\mathbf{x}) = ||\mathbf{x}||_p = \max\limits_{||\mathbf{z}||_q \leq 1} \mathbf{z}^T\mathbf{x}$ where q is such that $1/p + 1/q = 1$. Then
  $\partial f(\mathbf{x}) = \left\{ \mathbf{y} : ||\mathbf{y}||_q \leq 1 \text{ and } \mathbf{y}^T x = \max\limits_{||\mathbf{z}||_q \leq 1} \mathbf{z}^T\mathbf{x} \right\}$
- Can we derive the sub-differential of $||x||_1$?

Subgradients for the 'Lasso' Problem in Machine Learning

We use Lasso ($\min_{\mathbf{x}} f(\mathbf{x})$) as an example to illustrate subgradients of affine composition:

$$f(\mathbf{x}) = \frac{1}{2}||\mathbf{y} - \mathbf{x}||^2 + \lambda||\mathbf{x}||_1$$

The subgradients of $f(\mathbf{x})$ are

Subgradients for the 'Lasso' Problem in Machine Learning

We use Lasso ($\min_{\mathbf{x}} \; f(\mathbf{x})$) as an example to illustrate subgradients of affine composition:

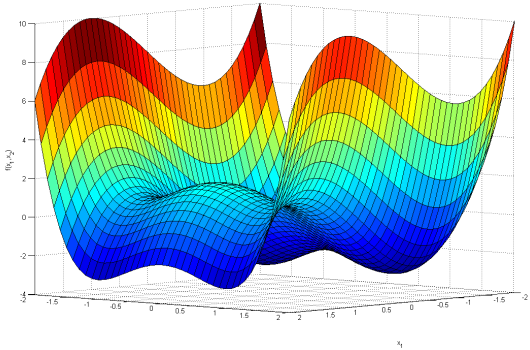
$$f(\mathbf{x}) = \frac{1}{2}||\mathbf{y} - \mathbf{x}||^2 + \lambda||\mathbf{x}||_1$$

The subgradients of $f(\mathbf{x})$ are

$$\mathbf{h} = \mathbf{x} - \mathbf{y} + \lambda\mathbf{s},$$

where $s_i = sign(x_i)$ if $x_i \neq 0$ and $s_i \in [-1, 1]$ if $x_i = 0$.

Figure below shows the plot of $f(x_1, x_2) = 3x_1^2 - x_1^3 - 2x_2^2 + x_2^4$. As can be seen in the plot, the function has several local maxima and minima.



More on Local Minima

- If a function f is differentiable, and x is a local minima, then $\nabla f(x) = 0$.

More on Local Minima

- If a function f is differentiable, and x is a local minima, then $\nabla f(x) = 0$.
- If f is not differentiable, then there counld be a local minima x with non-zero (sub)-gradient. Example: $f(x_1, x_2) = |x_1 - x_2|$. However, we can say that if x is a local minima, then $0 \in \partial f(x)$.

More on Local Minima

- If a function f is differentiable, and x is a local minima, then $\nabla f(x) = 0$.
- If f is not differentiable, then there counld be a local minima x with non-zero (sub)-gradient. Example: $f(x_1, x_2) = |x_1 - x_2|$. However, we can say that if x is a local minima, then $0 \in \partial f(x)$.
- Is the converse true? I.e. if x is s.t. $\nabla f(x) = 0$, then x is a local minima of f?

More on Local Minima

- If a function f is differentiable, and x is a local minima, then $\nabla f(x) = 0$.
- If f is not differentiable, then there counld be a local minima x with non-zero (sub)-gradient. Example: $f(x_1, x_2) = |x_1 - x_2|$. However, we can say that if x is a local minima, then $0 \in \partial f(x)$.
- Is the converse true? I.e. if x is s.t. $\nabla f(x) = 0$, then x is a local minima of f?
- No. For example, $f(x_1, x_2) = x_1^2 - x_2^2$. Such points are called saddle points!

Convexity and Global Minimum

Fundamental characteristics: **Let us now prove them**

1. Any point of local minimum point is also a point of global minimum.
2. For any stricly convex function, the point corresponding to the gobal minimum is also unique.

Convexity: Local and Global Minimum

Theorem

Let $f : \mathcal{D} \to \Re$ be a convex function on a convex domain $\mathcal{D}$. Any point of locally minimum solution for f is also a point of its globally minimum solution.

Proof: Suppose $\mathbf{x} \in \mathcal{D}$ is a point of local minimum and let $\mathbf{y} \in \mathcal{D}$ be a point of global minimum. Thus,

Convexity: Local and Global Minimum

Theorem

Let $f : \mathcal{D} \rightarrow \Re$ be a convex function on a convex domain $\mathcal{D}$. Any point of locally minimum solution for f is also a point of its globally minimum solution.

Proof: Suppose $\mathbf{x} \in \mathcal{D}$ is a point of local minimum and let $\mathbf{y} \in \mathcal{D}$ be a point of global minimum. Thus, $f(\mathbf{y}) < f(\mathbf{x})$. Since $\mathbf{x}$ corresponds to a local minimum, there exists an $\epsilon > 0$ such that

Convexity: Local and Global Minimum

Theorem

Let $f : \mathcal{D} \to \Re$ be a convex function on a convex domain $\mathcal{D}$. Any point of locally minimum solution for f is also a point of its globally minimum solution.

Proof: Suppose $\mathbf{x} \in \mathcal{D}$ is a point of local minimum and let $\mathbf{y} \in \mathcal{D}$ be a point of global minimum. Thus, $f(\mathbf{y}) < f(\mathbf{x})$. Since $\mathbf{x}$ corresponds to a local minimum, there exists an $\epsilon > 0$ such that

$$\forall\ \mathbf{z} \in \mathcal{D},\ ||\mathbf{z} - \mathbf{x}|| < \epsilon \Rightarrow f(\mathbf{z}) \geq f(\mathbf{x})$$

Consider a point $\mathbf{z}$

Convexity: Local and Global Minimum

Theorem

Let $f : \mathcal{D} \to \Re$ be a convex function on a convex domain $\mathcal{D}$. Any point of locally minimum solution for f is also a point of its globally minimum solution.

Proof: Suppose $\mathbf{x} \in \mathcal{D}$ is a point of local minimum and let $\mathbf{y} \in \mathcal{D}$ be a point of global minimum. Thus, $f(\mathbf{y}) < f(\mathbf{x})$. Since $\mathbf{x}$ corresponds to a local minimum, there exists an $\epsilon > 0$ such that

$$\forall\ \mathbf{z} \in \mathcal{D},\ ||\mathbf{z} - \mathbf{x}|| < \epsilon \Rightarrow f(\mathbf{z}) \geq f(\mathbf{x})$$

Consider a point $\mathbf{z} = \theta\mathbf{y} + (1 - \theta)\mathbf{x}$ with $\theta = \frac{\epsilon}{2||\mathbf{y} - \mathbf{x}||}$. Since $\mathbf{x}$ is a point of local minimum (in a ball of radius ϵ), and since $f(\mathbf{y}) < f(\mathbf{x})$, it must be that

Convexity: Local and Global Minimum

Theorem

Let $f : \mathcal{D} \to \Re$ be a convex function on a convex domain $\mathcal{D}$. Any point of locally minimum solution for f is also a point of its globally minimum solution.

Proof: Suppose $\mathbf{x} \in \mathcal{D}$ is a point of local minimum and let $\mathbf{y} \in \mathcal{D}$ be a point of global minimum. Thus, $f(\mathbf{y}) < f(\mathbf{x})$. Since $\mathbf{x}$ corresponds to a local minimum, there exists an $\epsilon > 0$ such that

$$\forall\ \mathbf{z} \in \mathcal{D},\ ||\mathbf{z} - \mathbf{x}|| < \epsilon \Rightarrow f(\mathbf{z}) \geq f(\mathbf{x})$$

Consider a point $\mathbf{z} = \theta \mathbf{y} + (1 - \theta)\mathbf{x}$ with $\theta = \frac{\epsilon}{2||\mathbf{y} - \mathbf{x}||}$. Since $\mathbf{x}$ is a point of local minimum (in a ball of radius ϵ), and since $f(\mathbf{y}) < f(\mathbf{x})$, it must be that $||\mathbf{y} - \mathbf{x}|| > \epsilon$. Thus, $0 < \theta < \frac{1}{2}$ and $\mathbf{z} \in \mathcal{D}$. Furthermore, $||\mathbf{z} - \mathbf{x}|| = \frac{\epsilon}{2}$.

Convexity: Local and Global Minimum (contd.)

Since f is a convex function

Convexity: Local and Global Minimum (contd.)

Since f is a convex function

$$f(\mathbf{z}) \leq \theta f(\mathbf{x}) + (1 - \theta) f(\mathbf{y})$$

Since $f(\mathbf{y}) < f(\mathbf{x})$, we also have

Convexity: Local and Global Minimum (contd.)

Since f is a convex function

$$f(\mathbf{z}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})$$

Since $f(\mathbf{y}) < f(\mathbf{x})$, we also have

$$\theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) < f(\mathbf{x})$$

The two equations imply that

Convexity: Local and Global Minimum (contd.)

Since f is a convex function

$$f(\mathbf{z}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})$$

Since $f(\mathbf{y}) < f(\mathbf{x})$, we also have

$$\theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) < f(\mathbf{x})$$

The two equations imply that $f(\mathbf{z}) < f(\mathbf{x})$, which contradicts our assumption that $\mathbf{x}$ corresponds to a point of local minimum. That is f cannot have a point of local minimum, which does not coincide with the point $\mathbf{y}$ of global minimum. $\square$

Since any locally minimum point for a convex function also corresponds to its global minimum, we will drop the qualifiers 'locally' as well as 'globally' while referring to the points corresponding to minimum values of a convex function.

Strict Convexity and Uniqueness of Global Minimum

For any stricly convex function, the point corresponding to the gobal minimum is also unique, as stated in the following theorem.

Theorem

Let $f : \mathcal{D} \rightarrow \Re$ be a strictly convex function on a convex domain $\mathcal{D}$. Then f has a unique point corresponding to its global minimum.

Proof: Suppose $\mathbf{x} \in \mathcal{D}$ and $\mathbf{y} \in \mathcal{D}$ with $\mathbf{y} \neq \mathbf{x}$ are two points of global minimum. That is $f(\mathbf{x}) = f(\mathbf{y})$ for $\mathbf{y} \neq \mathbf{x}$. The point $\frac{\mathbf{x}+\mathbf{y}}{2}$ also

Strict Convexity and Uniqueness of Global Minimum

For any stricly convex function, the point corresponding to the gobal minimum is also unique, as stated in the following theorem.

Theorem

Let $f : \mathcal{D} \to \Re$ be a strictly convex function on a convex domain $\mathcal{D}$. Then f has a unique point corresponding to its global minimum.

Proof: Suppose $\mathbf{x} \in \mathcal{D}$ and $\mathbf{y} \in \mathcal{D}$ with $\mathbf{y} \neq \mathbf{x}$ are two points of global minimum. That is $f(\mathbf{x}) = f(\mathbf{y})$ for $\mathbf{y} \neq \mathbf{x}$. The point $\frac{\mathbf{x}+\mathbf{y}}{2}$ also belongs to the convex set $\mathcal{D}$ and since f is strictly convex, we must have

$$f\left(\frac{\mathbf{x}+\mathbf{y}}{2}\right) < \frac{1}{2}f(\mathbf{x}) + \frac{1}{2}f(\mathbf{y}) = f(\mathbf{x})$$

which is a contradiction. Thus, the point corresponding to the minimum of f must be unique.

Does Global Minima Always Exist?

- Does the global minimum always exist?

Does Global Minima Always Exist?

- Does the global minimum always exist?
- Not necessarily even if f is bounded from below (e.g. $f(x) = e^x$)

Does Global Minima Always Exist?

- Does the global minimum always exist?
- Not necessarily even if f is bounded from below (e.g. $f(x) = e^x$)
- Weierstrass Theorem: Let f be a convex function and suppose there is a nonempty and bounded sublevel set $L_\alpha(f)$. Then f has a global minima.

Does Global Minima Always Exist?

- Does the global minimum always exist?
- Not necessarily even if f is bounded from below (e.g. $f(x) = e^x$)
- Weierstrass Theorem: Let f be a convex function and suppose there is a nonempty and bounded sublevel set $L_\alpha(f)$. Then f has a global minima.
- Since f is continuous, it attains a minimum over a closed and bounded ($=$ compact) set $L_\alpha(f)$ at some x^*. Note that x^* is also a global minimum as firstly, $f(x^*) \leq f(x), \forall x \in L_\alpha(f)$. Next since, $f(x^*) \leq \alpha$, it follows that for any $x \notin L_\alpha(f)$, $f(x) > \alpha \geq f(x^*)$

Critical Points are Global Minima for Convex Functions

- Lemma: Suppose that f is convex and differentiable over an open domain $dom(f)$. Let $x \in dom(f)$. Then if $\nabla f(x) = 0$ (i.e. a critical point), then x is a global minima.

Critical Points are Global Minima for Convex Functions

- Lemma: Suppose that f is convex and differentiable over an open domain $dom(f)$. Let $x \in dom(f)$. Then if $\nabla f(x) = 0$ (i.e. a critical point), then x is a global minima.
- Proof: Suppose $\nabla f(x) = 0$. Then from the first order characterization of convex functions, $\forall y \in dom(f), f(y) \geq f(x) + \nabla f(x)^T(y - x) \geq f(x)$. Hence x is a global minima.

Critical Points are Global Minima for Convex Functions

- Lemma: Suppose that f is convex and differentiable over an open domain $dom(f)$. Let $x \in dom(f)$. Then if $\nabla f(x) = 0$ (i.e. a critical point), then x is a global minima.
- Proof: Suppose $\nabla f(x) = 0$. Then from the first order characterization of convex functions, $\forall y \in dom(f), f(y) \geq f(x) + \nabla f(x)^T(y - x) \geq f(x)$. Hence x is a global minima.
- Note that this cannot be extended to non-differentiable convex functions since the global minima may not be a differentiable point (for example: $f(x) = ||x||_1$).

Critical Points are Global Minima for Convex Functions

- Lemma: Suppose that f is convex and differentiable over an open domain $dom(f)$. Let $x \in dom(f)$. Then if $\nabla f(x) = 0$ (i.e. a critical point), then x is a global minima.
- Proof: Suppose $\nabla f(x) = 0$. Then from the first order characterization of convex functions, $\forall y \in dom(f), f(y) \geq f(x) + \nabla f(x)^T(y - x) \geq f(x)$. Hence x is a global minima.
- Note that this cannot be extended to non-differentiable convex functions since the global minima may not be a differentiable point (for example: $f(x) = ||x||_1$).
- No Saddle points for convex functions!

Convex Optimization Problem

- Formally, a convex optimization problem is an optimization problem of the form

$$\begin{aligned} \textit{minimize } & f(\mathbf{w}) \\ \textit{subject to } & c \in C \end{aligned}$$

where f is a convex function, X is a convex set, and $\mathbf{w}$ is the optimization variable.

Convex Optimization Problem

- Formally, a convex optimization problem is an optimization problem of the form

$$\begin{aligned} \textit{minimize } & f(\mathbf{w}) \\ \textit{subject to } & c \in C \end{aligned}$$

  where f is a convex function, X is a convex set, and $\mathbf{w}$ is the optimization variable.
- if $X = dom(f)$, this becomes unconstrained optimization.

Convex Optimization Problem

- Formally, a convex optimization problem is an optimization problem of the form

$$\begin{aligned} minimize\ f(\mathbf{w}) \\ subject\ to\ c \ \in \ C \end{aligned}$$

  where f is a convex function, X is a convex set, and $\mathbf{w}$ is the optimization variable.
- if $X = dom(f)$, this becomes unconstrained optimization.
- A special case (f is a convex function, g_i are convex functions, and h_i are affine functions, and $\mathbf{x}$ is the vector of optimization variables):

$$\begin{aligned} minimize\ f(\mathbf{w}) & \\ subject\ to\ g_i(\mathbf{w}) & \leq 0,\ i = 1, ..., m \\ h_i(\mathbf{w}) & = 0, i = 1, ..., p \end{aligned}$$

Optimality Conditions for Constrained Optimization

- Lemma: Suppose that f is convex and differentiable over an open domain $dom(f)$. Let $X \subseteq dom(f)$ be a convex set. A point x^* is a minimizer of f over X if and only if

$$\nabla f(x^*)^T (x - x^*) \geq 0, \forall x \in X$$

Optimality Conditions for Constrained Optimization

- Lemma: Suppose that f is convex and differentiable over an open domain $dom(f)$. Let $X \subseteq dom(f)$ be a convex set. A point x^* is a minimizer of f over X if and only if

$$\nabla f(x^*)^T(x - x^*) \geq 0, \forall x \in X$$

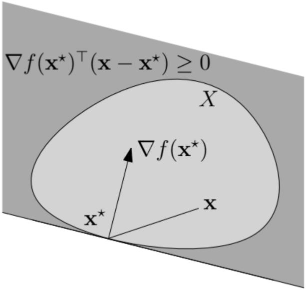
- Note that the Condition for Unconstrained minimization becomes a special case.

Optimality Conditions for Constrained Optimization

- Lemma: Suppose that f is convex and differentiable over an open domain $dom(f)$. Let $X \subseteq dom(f)$ be a convex set. A point x^* is a minimizer of f over X if and only if

$$\nabla f(x^*)^T(x - x^*) \geq 0, \forall x \in X$$

- Note that the Condition for Unconstrained minimization becomes a special case.
- Nice geometric interpretation:



Linear and Quadratic Programs

- Linear Program (LP) is a special case of a convex optimization problem:

$$\begin{aligned} \text{minimize } & c^T x \\ \text{subject to } & Ax \leq b \end{aligned}$$

Linear and Quadratic Programs

- Linear Program (LP) is a special case of a convex optimization problem:

$$\begin{aligned} \textit{minimize } & c^T x \\ \textit{subject to } & Ax \leq b \end{aligned}$$

- Another special case is Quadratic Programs (QP):

$$\begin{aligned} \textit{minimize } & 1/2 x^T Q x \\ \textit{subject to } & Ax \leq b \end{aligned}$$

Linear and Quadratic Programs

- Linear Program (LP) is a special case of a convex optimization problem:

$$\begin{aligned} &\mathit{minimize}\ c^T x \\ &\mathit{subject\ to}\ Ax \leq b \end{aligned}$$

- Another special case is Quadratic Programs (QP):

$$\begin{aligned} &\mathit{minimize}\ 1/2x^T Qx \\ &\mathit{subject\ to}\ Ax \leq b \end{aligned}$$

- The QP is a convex optimization problem only if Q is positive semi-definite,