CS6301: Optimization in Machine Learning

Lecture 5: Convexity and Convex Optimization Continued

Rishabh Iyer

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

January 22, 2020

Project and Assignment

- Project Deadline 1: Finalize on your Project Topics and partners: **February 15th 2020**
- Projects can be done in Groups with 1-3 students per group
- Project Deadline 2: Mid Term Review of the Project: **March 15th 2020**
- Final Project Report Deadline: **April 20th 2020**
- Last 3-5 Lectures of this class will be the course project presentations. Around 10 mins per project.
- **Updated Assignment Posted on eLearning. Due Date now is 5th February**

Outline

- Recap from Previous Lecture
- First order and second order properties of convex functions
- Basic Subgradient Calculus: Subgradients for non-differentiable convex functions
- Convex Optimization Problems
- Local and Global Minima of Convex Functions

Recap From Previous Lecture

- Convex Sets
- Definition of Convex Functions
- Convexity in 1 Dimension
- Definition of Convexity and Equivalent characterization in 1 Dimension
- Examples of Convex Functions
- Properties of Convex Functions and proving that functions are convex.

Convex Functions

- A Function $f : \mathbb{R}^d \to \mathbb{R}$ 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.

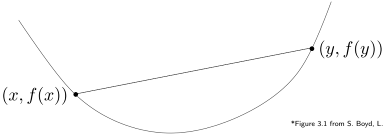
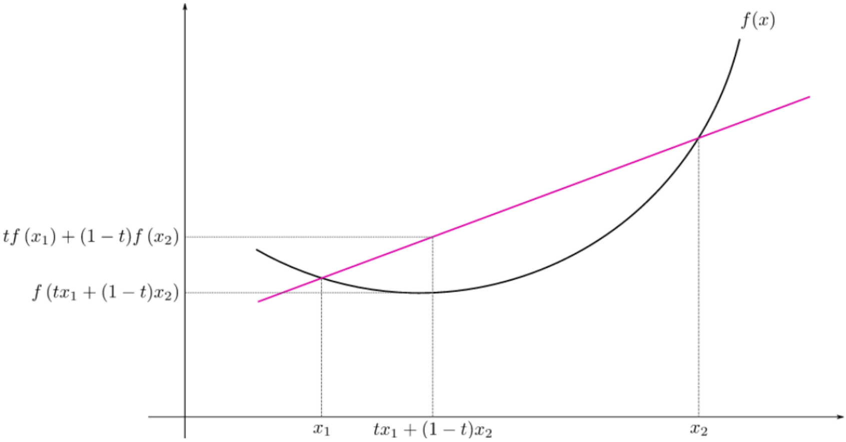


*Figure 3.1 from S. Boyd, L. Vandenberghe

- 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



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

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

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

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.

However a function f for which all its sublevel sets $L_\alpha(f)$ are convex sets, is not necessarily convex!

Such functions are called quasi convex

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

$$epi(f) = \{(\mathbf{x}, \alpha)|f(\mathbf{x}) \leq \alpha,\ \mathbf{x} \in \mathcal{D},\ \alpha \in \Re\} \quad (2)$$

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 f is convex if and only if $epi(f)$ is a convex set.

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* $\mathbf{x} \neq \mathbf{y}$,

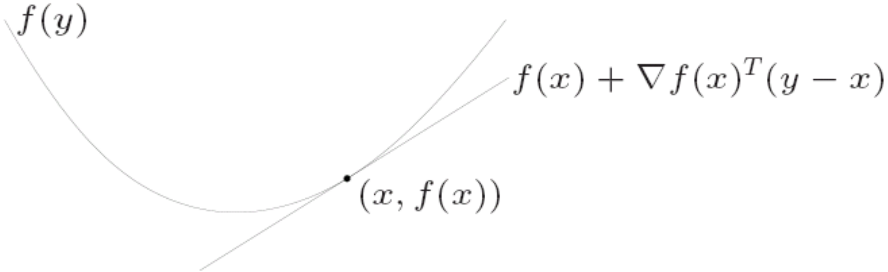
$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

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

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

Thus,

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

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

$$f(\mathbf{x}_2) = f(\mathbf{x}_1) + \nabla^T f(\mathbf{x}_1)(\mathbf{x}_2 - \mathbf{x}_1) \tag{3}$$

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

$$f(\theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2) = f(\mathbf{x}_2 + \theta(\mathbf{x}_1 - \mathbf{x}_2)) < \theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \quad (4)$$

Since (1) is already proved for convex functions, we use it in conjunction with (3), and (4), 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 $f''(x) \geq 0$

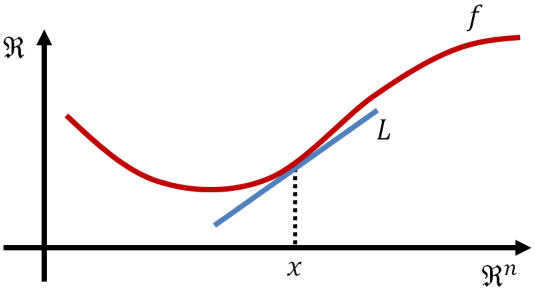
Monotonicity of Gradients

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!

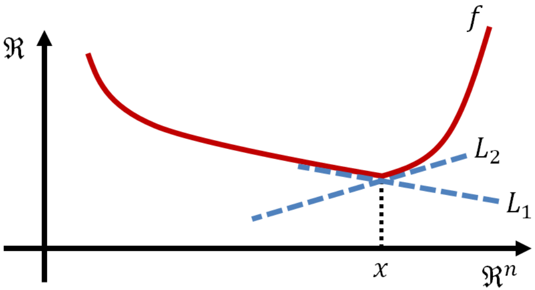
(Sub)Gradients and Convexity (contd)



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:

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

(Sub)Gradients and Convexity (contd)

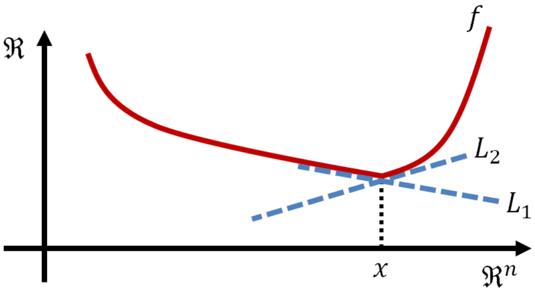


In this figure we see the function f at $\mathbf{x}$ has many possible linear tangents that may fit appropriately. Then a **subgradient** is any $\mathbf{h} \in \Re^n$ (same dimension as x) such that:

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

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 $f(\mathbf{x}) = max_{i=1...m} f_i(\mathbf{x})$, 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), $\partial f(\mathbf{x}) = conv\left(\bigcup_{s:f_s(\mathbf{x})=f(\mathbf{x})} \partial f_s(\mathbf{x})\right)$

Subgradient of $\max\{w^T x, 0\}$

Assume $x \in \Re^n$. Then

- This function is not differentiable on the hyper-plane $H = \{x | w^T x = 0\}$.

Subgradient of $\max\{w^T x, 0\}$

Assume $x \in \Re^n$. Then

- This function is not differentiable on the hyper-plane $H = \{x | w^T x = 0\}$.
- From the previous slide, we know that for any $x \in H$, $\partial f(x) = conv(w, 0)$.

Subgradient of $\max\{w^T x, 0\}$

Assume $x \in \Re^n$. Then

- This function is not differentiable on the hyper-plane $H = \{x | w^T x = 0\}$.
- From the previous slide, we know that for any $x \in H$, $\partial f(x) = conv(w, 0)$.
- In other words, $\partial f(x) = \{\lambda w\}$ for $\lambda \geq 0$.

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, $\partial\|\mathbf{x}\|_1 = conv\left(\bigcup_{\mathbf{s}\in\mathcal{S}^*} \mathbf{s}\right)$.
- In simple terms, the subgradient $h = s$, where s satisfies $s_i = sign(x_i)$ if $x_i \neq 0$ and $s_i \in [-1, 1]$ if $x_i = 0$.

More of Basic Subgradient Calculus

Here, we provide some basic subgradient calculus for convex functions:

- 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 $\partial g(\mathbf{x}) = A^T \partial f(A\mathbf{x} + b)$

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:

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

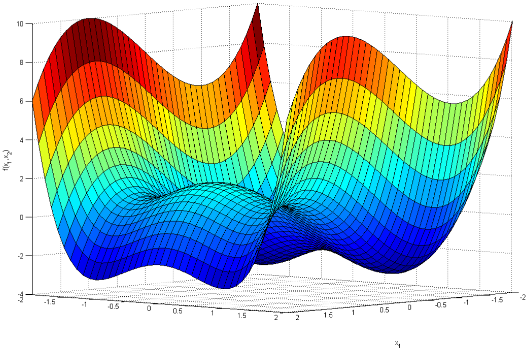
The subgradients of $f(\mathbf{x})$ are

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

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

Local Minima

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.



Definition of Local Minima

Definition

A Point $\mathbf{x}$ is a local minimum of a function f, if there exists an $\epsilon > 0$ such that

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

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 $f : \mathcal{D} \rightarrow \Re$ 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,

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

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

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

  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.
- 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):

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

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

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

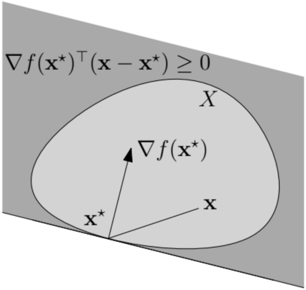
- Note that the Condition for Unconstrained minimization becomes a special case.

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

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

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):

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

Linear and Quadratic Programs

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

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

- Another special case is Quadratic Programs (QP):

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

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