Optimization in Machine Learning

Lecture 10: Necessary and sufficient conditions for optimization with and without Convexity, Lipschitz Continuity

Ganesh Ramakrishnan

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

February, 2025

Outline

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

More on SubGradient kind of functions: Monotonicity (contd)

Definition

2. $\mathbf{h}$ is *strictly monotone* on $\mathcal{D}$ if for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ with $\mathbf{x}_1 \neq \mathbf{x}_2$,

$$\left(\mathbf{h}(\mathbf{x}_1) - \mathbf{h}(\mathbf{x}_2)\right)^T (\mathbf{x}_1 - \mathbf{x}_2) > 0 \tag{1}$$

3. $\mathbf{h}$ is *uniformly* or *strongly monotone* on $\mathcal{D}$ if for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, there is a constant $c > 0$ such that

$$\left(\mathbf{h}(\mathbf{x}_1) - \mathbf{h}(\mathbf{x}_2)\right)^T (\mathbf{x}_1 - \mathbf{x}_2) \geq c||\mathbf{x}_1 - \mathbf{x}_2||^2 \tag{2}$$

(Sub)Gradients and Convexity

Relationship between convexity of a function and monotonicity of its (sub)gradient:

Theorem

Let $f : \mathcal{D} \to \Re$ with $\mathcal{D} \subseteq \Re^n$ be differentiable on the convex set $\mathcal{D}$. Then,

1. *f is convex on $\mathcal{D}$* **iff** *its gradient ∇f is monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$:* $\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) \geq 0$
2. *f is strictly convex on $\mathcal{D}$* **iff** *its gradient ∇f is strictly monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$ with $\mathbf{x} \neq \mathbf{y}$:* $\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) > 0$
3. *f is uniformly or strongly convex on $\mathcal{D}$* **iff** *its gradient ∇f is uniformly monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$,* $\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) \geq c||\mathbf{x} - \mathbf{y}||^2$ *for some constant $c > 0$.*

While these results also hold for (more advanced proximal) subgradients $\mathbf{h}_p$ (see `https://moodle.iitb.ac.in/mod/resource/view.php?id=32806`), we will quickly show them only for gradients ∇f

Advanced: $\mathbf{h}_p$ is a proximal gradient of f at $\mathbf{x}$ iff, $\forall \mathbf{y} \in dmn(f)$, $f(y) \geq f(x) + \mathbf{h}_p(\mathbf{y} - \mathbf{x}) - \frac{\lambda}{2}||\mathbf{y} - \mathbf{x}||^2$

(Sub)Gradients and Convexity (contd)

Proof:
Necessity: Suppose f is strongly convex on $\mathcal{D}$. Then we know from an earlier result that for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,

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

Adding the two inequalities, we get uniform/strong monotonicity in definition (3). If f is convex, the inequalities hold with $c = 0$, yielding monotonicity in definition (1). If f is strictly convex, the inequalities will be strict, yielding strict monotonicity in definition (2).

(Sub)Gradients and Convexity (contd)

Sufficiency: Suppose ∇f is monotone. For any fixed $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, consider the function $\phi(t) = f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))$. By the mean value theorem applied to $\phi(t)$, we should have for some $t \in (0, 1)$,

$$\phi(1) - \phi(0) = \phi'(t) \tag{3}$$

Letting $\mathbf{z} = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$, (3) translates to

$$f(\mathbf{y}) - f(\mathbf{x}) = \nabla^T f(\mathbf{z})(\mathbf{y} - \mathbf{x}) \tag{4}$$

Also, by definition of monotonicity of ∇f,

$$(\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x}) = \frac{1}{t} (\nabla f(\mathbf{z}) - \nabla f(\mathbf{x}))^T (\mathbf{z} - \mathbf{x}) \geq 0 \tag{5}$$

(Sub)Gradients and Convexity (contd)

Combining (4) with (5), we get,

$$f(\mathbf{y}) - f(\mathbf{x}) = (\nabla f(\mathbf{z}) - f(\mathbf{x}))^T (\mathbf{y} - \mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \\ \geq \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \tag{6}$$

By a previous foundational result, this inequality proves that f is convex. Strict convexity can be similarly proved by using the strict inequality in (5) inherited from strict monotonicity, and letting the strict inequality follow through to (6).

(Sub)Gradients and Convexity (contd)

For the case of strong convexity, we have

$$\phi'(t) - \phi'(0) = \left(\nabla f(\mathbf{z}) - f(\mathbf{x})\right)^T (\mathbf{y} - \mathbf{x})$$
$$= \frac{1}{t}\left(\nabla f(\mathbf{z}) - f(\mathbf{x})\right)^T (\mathbf{z} - \mathbf{x}) \geq \frac{1}{t}c||\mathbf{z} - \mathbf{x}||^2 = ct||\mathbf{y} - \mathbf{x}||^2 \tag{7}$$

Therefore,

$$\phi(1) - \phi(0) - \phi'(0) = \int_0^1 [\phi'(t) - \phi'(0)]dt \geq \frac{1}{2}c||\mathbf{y} - \mathbf{x}||^2 \tag{8}$$

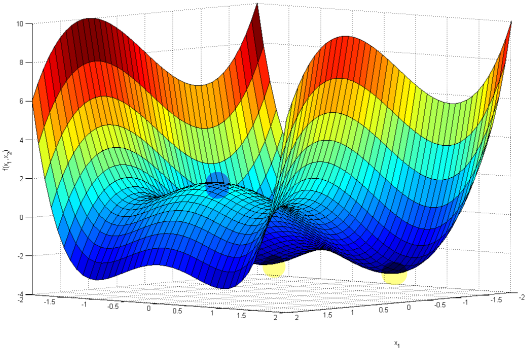
which translates to

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

Thus, f must be strongly convex.

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.



Formally: Maximum and Minimum values of functions

Let $f : \mathcal{D} \rightarrow \Re$. Now f has

- An *absolute maximum* (or global maximum) value at point $\mathbf{c} \in \mathcal{D}$ if

$$f(\mathbf{x}) \leq f(\mathbf{c}), \ \forall \mathbf{x} \in \mathcal{D}$$

- An *absolute minimum* (or global minimum) value at $\mathbf{c} \in \mathcal{D}$ if

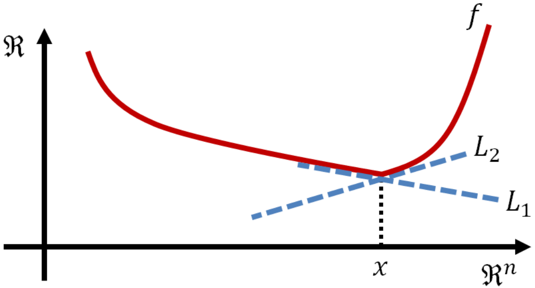
$$f(\mathbf{x}) \geq f(\mathbf{c}), \ \forall \mathbf{x} \in \mathcal{D}$$

- A *local maximum value* at $\mathbf{c}$ if there is an open ball $\mathcal{B}$ containing $\mathbf{c}$ in which $f(\mathbf{c}) \geq f(\mathbf{x}), \ \forall x \in \mathcal{I}$
- A *local minimum value* at c if there is an open ball $\mathcal{B}$ containing c in which $f(\mathbf{c}) \leq f(\mathbf{x}), \ \forall \mathbf{x} \in \mathcal{B}$
- A *local extreme value* at $\mathbf{c}$, if $f(\mathbf{c})$ is either a local maximum or local minimum value of f in an open ball $\mathcal{B}$ with $c \in \mathcal{B}$

Convexity, Subgradients and Minima

1. Subgradient-based sufficient condition for global minimum
2. Any point of local minimum point is also a point of global minimum.
3. For any stricly convex function, the point corresponding to the global minimum is also unique.

(Sub)Gradients and Minima



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

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

What is its connection with Minima of the function?

(Sub)Gradients and Optimality: Sufficient Condition

- For a convex f,

(Sub)Gradients and Optimality: Sufficient Condition

- For a convex f,

$\forall\, y \in dmn(f) \qquad f(y) \geq f(x) + h_x^T(y-x)$

If at x there exists $0 \in \partial f(x)$

ie $h_x = 0$ is a valid subgradient

Then

$\forall\, y \in dmn(f) \qquad f(y) \geq f(x)$

(Sub)Gradients and Optimality: Sufficient Condition

- For a convex f,
$$f(\mathbf{x}^*) = \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \Leftarrow 0 \in \partial f(\mathbf{x}^*)$$
- The reason: $\mathbf{h} = 0$ being a subgradient means that for all $\mathbf{y}$

(Sub)Gradients and Optimality: Sufficient Condition

- For a convex f,

$$f(\mathbf{x}^*) = \min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \Leftarrow 0 \in \partial f(\mathbf{x}^*)$$

- The reason: $\mathbf{h} = 0$ being a subgradient means that for all $\mathbf{y}$

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) + 0^T(\mathbf{y} - \mathbf{x}^*) = f(\mathbf{x}^*)$$

- The analogy to the differentiable case is: $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$.
- Thus, for a convex function $f(\mathbf{x})$, if $\nabla f(\mathbf{x}) = 0$, then $\mathbf{x}$ must be a point of global minimum.
- Is there a necessary condition for a differentiable (possibly non-convex) function having a (local or global) minimum at $\mathbf{x}$? (We will see broadly)

Critical Points of convex functions are Global Minima: Sufficient Condition

Corollary of the more general result on the previous slide is...

Critical Points of convex functions are Global Minima: Sufficient Condition

Corollary of the more general result on the previous slide is...

- Corolloary: 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 minimum.
- 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 minimum.
- Note that the extension to global minimum of non-differentiable convex functions is precisely through the 0 subgradient $\mathbf{h} = 0$ (discussed on the previous slide)
  - Recap from last class, we derived the global minimum for $f(\mathbf{x}) = \frac{1}{2}||\mathbf{y} - \mathbf{x}||^2 + \lambda||\mathbf{x}||_1$
- No saddle point for convex functions!
  - $\mathbf{x}$ is called a saddle point of f if $\nabla f(\mathbf{x}) = 0$ but $\mathbf{x}$ is neither a local maximum nor a local minimum!

Illustration of Saddle point: Impossible for Convex Function

As an example, the (non-convex) function $f(x) = 3x^5 - 5x^3$ has the derivative $f'(x) = 15x^2(x+1)(x-1)$. The critical points are 0, 1 and -1. Of the three, the sign of $f'(x)$ changes at 1 and -1, which are local minimum and maximum respectively. The sign does not change at 0, which is therefore not a local supremum.

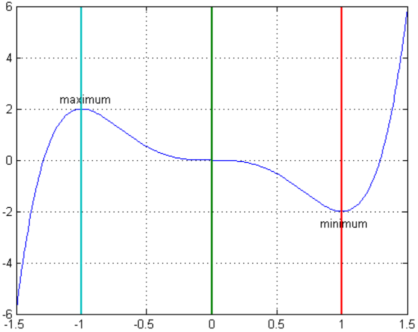
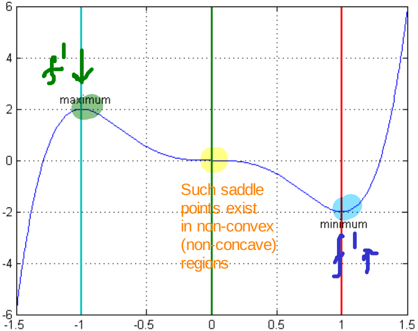


Illustration of Saddle point: Impossible for Convex Function

As an example, the (non-convex) function $f(x) = 3x^5 - 5x^3$ has the derivative $f'(x) = 15x^2(x+1)(x-1)$. The critical points are 0, 1 and -1. Of the three, the sign of $f'(x)$ changes at 1 and -1, which are local minimum and maximum respectively. The sign does not change at 0, which is therefore not a local supremum.



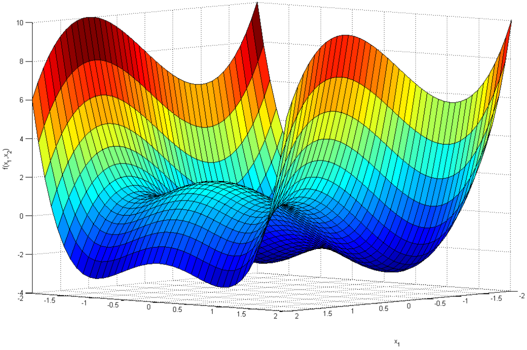
[Exercise] Illustration of Saddle Point

Figure below shows the plot of a non-convex $f(x_1, x_2) = 3x_1^2 - x_1^3 - 2x_2^2 + x_2^4$. Can we identify its saddle points?

Lots of analysis in the deep neural network regime is about saddle point optimization

Simpler fn

$x_1^2 - x_2^2$



Subgradients in Lasso: Sufficient Condition Test

We illustrate the sufficient condition again using Lasso as an example. Consider the simplified Lasso problem:

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

Recall the subgradients of $f(\mathbf{x})$:

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

where $s_i = sign(x_i)$ if $x_i \neq 0$ and $s_i \in [-1, 1]$ if $x_i = 0$.
A solution to this problem is

Subgradients in Lasso: Sufficient Condition Test

We illustrate the sufficient condition again using Lasso as an example. Consider the simplified Lasso problem:

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

Recall the subgradients of $f(\mathbf{x})$:

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

Case 1 Case 2

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

convex hull for subdifferential

A solution to this problem is

Based on a sufficient condition from the previous slide

Case 1: Shrink beyond closed interval

If $h_i = 0$ & $x_i > 0 \Leftrightarrow h_i = x_i - y_i + \lambda = 0 \Leftrightarrow x_i = y_i - \lambda > 0 \Leftrightarrow x_i = y_i - \lambda$ if $y_i > \lambda$

If $h_i = 0$ & $x_i < 0 \Leftrightarrow h_i = x_i - y_i - \lambda = 0 \Leftrightarrow x_i = y_i + \lambda < 0 \Leftrightarrow x_i = y_i + \lambda$ if $y_i < -\lambda$

CASE 2?

$h_i = 0$ & $x_i = 0 \Leftrightarrow h_i = -y_i + \lambda s_i = 0 \Leftrightarrow$ $x_i = 0$ if $y_i \in [-\lambda, \lambda]$

$s_i \in [-1, +1]$

Subgradients in Lasso: Sufficient Condition Test

We illustrate the sufficient condition again using Lasso as an example. Consider the simplified Lasso problem:

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

Recall the subgradients of $f(\mathbf{x})$:

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

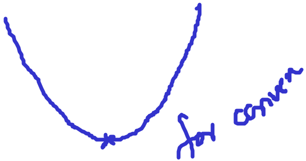
where $s_i = sign(x_i)$ if $x_i \neq 0$ and $s_i \in [-1, 1]$ if $x_i = 0$.
A solution to this problem is $\mathbf{x}^* = S_\lambda(\mathbf{y})$, where $S_\lambda(\mathbf{y})$ is the soft-thresholding operator:

$$S_\lambda(\mathbf{y}) = \begin{cases} y_i - \lambda & \text{if } y_i > \lambda \\ 0 & \text{if } -\lambda \leq y_i \leq \lambda \\ y_i + \lambda & \text{if } y_i < -\lambda \end{cases}$$

Now let $\mathbf{x}^* = S_\lambda(\mathbf{y})$ and we can get $\mathbf{h} = 0$. Why? If $y_i > \lambda$, we have $x_i^* - y_i = -\lambda + \lambda \cdot 1 = 0$. The case of $y_i < \lambda$ is similar. If $-\lambda \leq y_i \leq \lambda$, we have $x_i^* - y_i = -y_i + \lambda(\frac{y_i}{\lambda}) = 0$. Here, $s_i = \frac{y_i}{\lambda}$.

Convexity, Subgradients and Minima

1. Subgradient-based sufficient condition for global minimum [Done]
2. Any point of local minimum point is also a point of global minimum. **[Let us now prove it next]**
3. For any stricly convex function, the point corresponding to the global 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,

Theorem

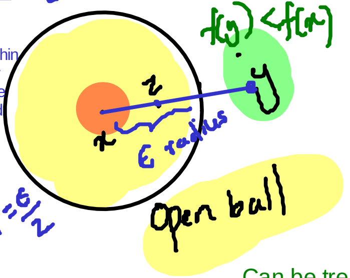
Let $f : \mathcal{D} \rightarrow \Re$ be a convex function on a convex domain $\mathcal{D}$. Any point of locally minimum solution for f is also a point of its globally minimum solution.

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,

$\exists$ a ball B of radius ϵ around x

Idea behind the proof:
Construct a point z within the open call and show
f(z) < f(x) just because
f(y) < f(x) was claimed

$\|z-x\| = \epsilon/2$



for f(x)=f(y)

Can be treated as a proof by contradiction

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

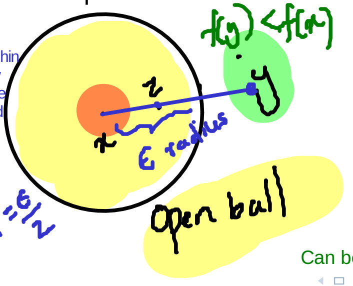
Convexity: Local and Global Minimum

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

Idea behind the proof:
Construct a point z within
the open ball and show
f(z) < f(x) just because
f(y) < f(x) was claimed



$\|z-x\| = \epsilon/2$

Can be treated as a proof by contradiction

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

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

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

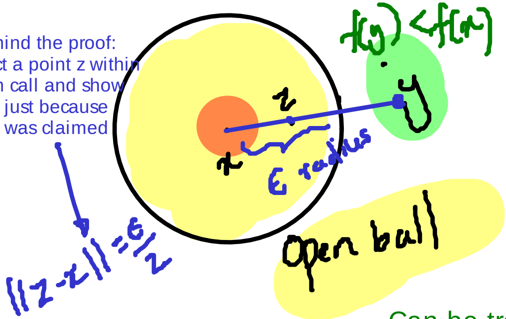
$$\forall\, \mathbf{z} \in \mathcal{D},\ ||\mathbf{z} - \mathbf{x}|| < \epsilon \Rightarrow f(\mathbf{z}) \geq f(\mathbf{x})$$

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

Idea behind the proof:
Construct a point z within the open call and show f(z) < f(x) just because f(y) < f(x) was claimed



Can be treated as a proof by contradiction

Convexity: Local and Global Minimum (contd.)

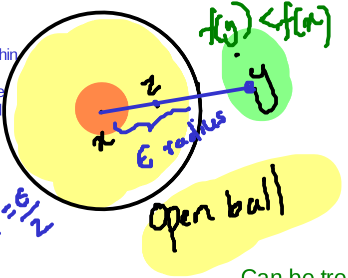
Since f is a convex function

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

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

Idea behind the proof:
Construct a point z within
the open call and show
f(z) < f(x) just because
f(y) < f(x) was claimed

$\|z-x\| = \epsilon/2$



Can be treated as a proof by contradiction

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. □

As we will, since any locally minimum point for a convex function also corresponds to its global minimum, we will soon 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 $f : \mathcal{D} \to \Re$ 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

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

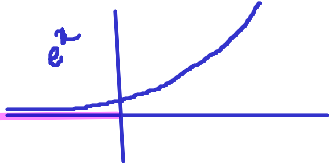
$$f\left(\frac{\mathbf{x}+\mathbf{y}}{2}\right) < \frac{1}{2}f(\mathbf{x}) + \frac{1}{2}f(\mathbf{y}) = f(\mathbf{x})$$

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

Does Global Minimum Always Exist?

- Does the global minimum always exist?
- Not necessarily even if f is bounded from below
  - *E.g.*, $f(x) = e^x$ is bounded below by 0, but the bound is never achieved
- 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 minimum.
- 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 $x \notin L_\alpha(f)$, $f(x) > \alpha \geq f(x^*)$

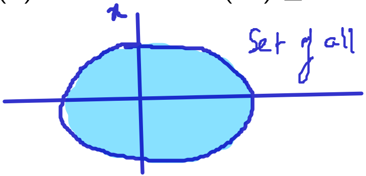
Does Global Minimum Always Exist?



- Does the global minimum always exist?
- Not necessarily even if f is bounded from below
  - *E.g.*, $f(x) = e^x$ is bounded below by 0, but the bound is never achieved
- 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 minimum.

If the function has a strict unique global minimum and the function is differentiable (in addition to convexity) can we claim that the reverse also holds? Appears YES

- 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 $x \notin L_\alpha(f)$, $f(x) > \alpha \geq f(x^*)$



Set of all pts $L_\alpha(f)$

$\{x \mid f(x) \leq \alpha\}$

Critical Points are Global Minima

- 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: $f(x) = ||x||_1$).

Convex Optimization Problem

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

$$\begin{aligned} \textit{minimize } & f(\mathbf{x}) \\ \textit{subject to } & \mathbf{x} \in \mathcal{X} \end{aligned}$$

  where f is a convex function, $\mathcal{X}$ is a convex set, and $\mathbf{x}$ is the optimization variable.
- if $\mathcal{X} = dom(f)$, 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):

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

Convex Optimization Problem

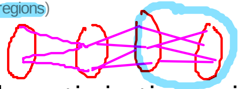
- Formally, a convex optimization problem is an optimization problem of the form

if $x^* \in int(\mathcal{X})$ then $0 \in \partial f(x^*)$ is sufficient

But the convexity of the set (convexity of the constraint set) is typically not relaxed

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) \\ &\text{subject to } \mathbf{x} \in \mathcal{X} \end{aligned}$$

Relaxations are typically on the function f (need not be convex everywhere though it might be convex in some regions)

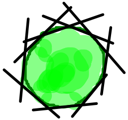


where f is a convex function, $\mathcal{X}$ is a convex set, and $\mathbf{x}$ is the optimization variable.

- if $\mathcal{X} = dom(f)$, this becomes unconstrained optimization.

Eg: Region could be defined as the region of last layer weights for a deep NN (that with all other layer weights frozen)

- 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):



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

Intersection of convex 0-sublevel sets of g_i's

Typical convex set. In fact the dual description for convex sets should allow them to be specified as intersections of such inequalities

Either or both of m and p could be 0