Convexity, Local and Global Optimality, etc.
Recap: Some Interesting Connections in $\mathbb{R}^n$

1. The closure of a set is the smallest closed set containing the set. The closure of a closed set is the set itself.

2. $S$ is closed if and only if $\text{closure}(S) = S$.

3. A bounded set can be defined in terms of a closed set; A set $S$ is bounded if and only if it is contained strictly inside a closed set.

4. A relationship between the interior, boundary and closure of a set $S$ is $\text{closure}(S) = \text{int}(S) \cup \partial(S)$. 
Extending Open, Closed sets, Boundary, Interior, etc to Topological Sets

This is for Optimal Reading

1 Recap: Open Set follows from Definition 1 of Topology. Neighborhood follows from Definition 2 of Topology. **By this definition, can point in interior be limit point?**

2 **Limit Point:** Let \( S \) be a subset of a topological set \( X \). A point \( x \in X \) is a limit point of \( S \) if every neighborhood of \( x \) contains at least one point of \( S \) different from \( x \) itself.
   - If \( X \) has an associated metric \( d \) and \( S \subseteq X \) then \( x \in S \) is a limit point of \( S \) if \( \forall \epsilon > 0, \{y \in S \mid 0 < d(y, x) < \epsilon\} \neq \emptyset \).

3 **Closure of \( S \)** = \( \text{closure}(S) = S \cup \{\text{limit points of } S\} \).

4 **Boundary \( \partial S \) of \( S \):** Is the subset of \( S \) such that every neighborhood of a point from \( \partial S \) contains at least one point in \( S \) and one point not in \( S \).
   - If \( S \) has a metric \( d \) then:
     \[
     \partial S = \{x \in S \mid \forall \epsilon > 0, \exists \ y \text{ s.t. } d(x, y) < \epsilon \text{ and } y \in S \text{ and } \exists \ z \text{ s.t. } d(x, z) < \epsilon \text{ and } z \notin S \}
     \]

5 **Open set \( S \):** Does not contain any of its boundary points
   - If \( X \) has an associated metric \( d \) and \( S \subseteq X \) is called open if for any \( x \in S \), \( \exists \ \epsilon > 0 \) such that given any \( y \in S \) with \( d(y, x) < \epsilon \), \( y \in S \).

6 **Closed set \( S \):** Has an open complement \( S^C \)
Revisiting Example for Local Extrema

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

Figure 1: A local min
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 strictly convex function, the point corresponding to the global minimum is also unique.
Convexity: Local and Global Minimum

Theorem

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

Proof: Suppose $x \in D$ is a point of local minimum and let $y \in D$ be a point of global minimum. Thus, $f(y) < f(x)$.

We are trying to prove by contradiction that a $y$ different from $x$ cannot exist.
**Theorem**

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

**Proof:** Suppose $x \in D$ is a point of local minimum and let $y \in D$ be a point of global minimum. Thus, $f(y) < f(x)$. Since $x$ corresponds to a local minimum, there exists an $\epsilon > 0$ such that for all points in the epsilon disc, the value is $\geq f(x)$.
Convexity: Local and Global Minimum

Theorem

Let \( f : D \rightarrow \mathbb{R} \) be a convex function on a convex domain \( D \). Any point of locally minimum solution for \( f \) is also a point of its globally minimum solution.

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

\[
\forall z \in D, \ |x - z| < \epsilon \Rightarrow f(z) \geq f(x)
\]

Consider a point \( z \) lying on the line segment joining \( x \) and \( y \) but lying inside the epsilon disc. We show that \( f(z) < f(x) \) contradicting the assumption that \( x \) was a local min in the epsilon disc.
Convexity: Local and Global Minimum

**Theorem**

Let \( f : D \to \mathbb{R} \) be a convex function on a convex domain \( D \). Any point of locally minimum solution for \( f \) is also a point of its globally minimum solution.

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

\[
\forall \ z \in D, \quad ||z - x|| < \epsilon \Rightarrow f(z) \geq f(x)
\]

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

We have shown a specific value for theta when we assume a norm
Convexity: Local and Global Minimum

**Theorem**

Let \( f : D \to \mathbb{R} \) be a convex function on a convex domain \( D \). Any point of locally minimum solution for \( f \) is also a point of its globally minimum solution.

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

\[
\forall z \in D, \ |z - x| < \epsilon \Rightarrow f(z) \geq f(x)
\]

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

Since $f$ is a convex function

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

Since $f(y) < f(x)$, we also have

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

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

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 strictly convex function, the point corresponding to the global minimum is also unique, as stated in the following theorem.

**Theorem**

Let \( f : D \rightarrow \mathbb{R} \) be a strictly convex function on a convex domain \( D \). Then \( f \) has a unique point corresponding to its global minimum.

**Proof:** Suppose \( x \in D \) and \( y \in D \) with \( y \neq x \) are two points of global minimum. That is \( f(x) = f(y) \) for \( y \neq x \). The point \( \frac{x+y}{2} \) also should lie in \( D \).
Strict Convexity and Uniqueness of Global Minimum

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

**Theorem**

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

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

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

which is a contradiction. Thus, the point corresponding to the minimum of $f$ must be unique.
It is possible that a convex function is NOT strictly convex and yet it has a unique global minimum.
Recap for differentiable $f: \mathbb{R} \rightarrow \mathbb{R}$ the equivalent definition of convexity:

A nondecreasing $f'$
Convexity and Differentiability

1. Recap for differentiable $f : \mathbb{R} \rightarrow \mathbb{R}$ the equivalent definition of convexity
2. What would be an equivalent notion of differentiability and convexity for $f : \mathbb{R}^n \rightarrow \mathbb{R}$?
3. What will be critical points? First and second order necessary (and sufficient) conditions for local and global optimality?

$$3x^2 - x + y^2$$
In both views, I find that the convexity of the function is reflected in the non-decreasing nature of the derivatives along the respective axis (directions).
How about convexity in an arbitrary direction?

Expect the directional derivative of the convex function to be non-decreasing along EVERY direction.

Is there a more compact mathematical expression for this?
Optimization Principles for Multivariate Functions

In the following, we state some important properties of convex functions, some of which require knowledge of ‘derivatives’ in $\mathbb{R}^n$. These also include relationships between convex functions and convex sets, and first and second order conditions for convexity.
The Direction Vector

- Consider a function $f(x)$, with $x \in \mathbb{R}^n$.
- We start with the concept of the direction at a point $x \in \mathbb{R}^n$.
- We will represent a vector by $x$ and the $k^{th}$ component of $x$ by $x_k$.
- Let $u^k$ be a unit vector pointing along the $k^{th}$ coordinate axis in $\mathbb{R}^n$.
- An arbitrary direction vector $v$ at $x$ is a vector in $\mathbb{R}^n$ with unit norm (i.e., $\|v\| = 1$) and component $v_k$ in the direction of $u^k$. 
Directional derivative and the gradient vector

Let \( f : D \rightarrow \mathbb{R}, \ D \subseteq \mathbb{R} \) be a function.

**Definition**

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

\[
D_v f(x) = \lim_{h \to 0} \frac{f(x + hv) - f(x)}{h}
\]

provided the limit exists.
Directional Derivative

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

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

Claim

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

$$D_v f(x) = \sum_{k=1}^{n} \frac{\partial f(x)}{\partial x_k} v_k$$ (2)
Directional Derivative: Simplified Expression

Define $g(h) = f(x + vh)$. Now:

- $g'(0) = f'(x+vh)$ evaluated at $h=0$

A more formal derivation of Directional derivative as dot product of gradient with vector $v$
Directional Derivative: Simplified Expression

Define $g(h) = f(x + vh)$. Now:

- $g'(0) = \lim_{h \to 0} \frac{g(0+h) - g(0)}{h} = \lim_{h \to 0} \frac{f(x + hv) - f(x)}{h}$, which is the expression for the directional derivative defined in equation 1. Thus, $g'(0) = D_v f(x)$.

- By definition of the chain rule for partial differentiation, we get another expression for $g'(0)$ as

$g'(0) = \boxed{\text{expression}}$
Directional Derivative: Simplified Expression

Define \( g(h) = f(x + vh) \). Now:

\[
g'(0) = \lim_{h \to 0} \frac{g(0+h) - g(0)}{h} = \lim_{h \to 0} \frac{f(x+vh) - f(x)}{h},
\]
which is the expression for the directional derivative defined in equation 1. Thus, \( g'(0) = D_v f(x) \).

By definition of the chain rule for partial differentiation, we get another expression for \( g'(0) \) as

\[
g'(0) = \sum_{k=1}^{n} \frac{\partial f(x)}{\partial x_k} v_k \sum_{k=1}^{n} \partial f(x)
\]

Therefore, \( g'(0) = D_v f(x) = \sum_{k=1}^{n} \frac{\partial f(x)}{\partial x_k} v_k \). \( \square \)

Homeworks:

1. Consider the polynomial \( f(x, y, z) = xy + z \sin xy \) and the unit vector \( v \triangleq \frac{1}{\sqrt{3}} [1, 1, 1] \). Consider the point \( p_0 = (0, 1, 3) \). Compute the directional derivative of \( f \) at \( p_0 \) in the direction of \( v \).

2. Compute the rate of change of \( f(x, y, z) = e^{xyz} \) at \( p_0 = (1, 2, 3) \) in the direction from \( p = (1, 2, 3) \) to \( p = (-4, 6, -1) \).
Illustrating Computation of Directional Derivative

Consider the polynomial \( f(x, y, z) = x \dot{y} + z \sin xy \) and the unit vector \( \mathbf{v} = \frac{2}{\sqrt{3}} [1, 1, 1]^T \). Consider the point \( p_0 = (0, 1, 3) \). We will compute the directional derivative of \( f \) at \( p_0 \) in the direction of \( \mathbf{v} \).
More on the Gradient Vector

- All our ideas about first and second derivative in the case of a single variable carry over to the directional derivative.

- What does the gradient $\nabla f(x)$ tell you about the function $f(x)$? While there exist infinitely many direction vectors $v$ at any point $x$, there is a unique gradient vector $\nabla f(x)$.

- Since we expressed $D_v f(x)$ as the dot product of $\nabla f(x)$ with $v$, we can study $\nabla f(x)$ independently.

The gradient vector as a canonical representation of the directional derivative but expressed independent of any direction needs some insight (geometrical as well)
More on the Gradient Vector

- All our ideas about first and second derivative in the case of a single variable carry over to the directional derivative.
- What does the gradient $\nabla f(x)$ tell you about the function $f(x)$? While there exist infinitely many direction vectors $v$ at any point $x$, there is a unique gradient vector $\nabla f(x)$.
- Since we expressed $D_vf(x)$ as the dot product of $\nabla f(x)$ with $v$, we can study $\nabla f(x)$ independently.

Claim

Suppose $f$ is a differentiable function of $x \in \mathbb{R}^n$. The maximum value of the directional derivative $D_vf(x)$ is $||\text{gradient of } f(x) ||$ assuming $v$ has unit L2 norm. Proof?

Will depend in general on the norm under which $v$ has a unit value.

Steepest descent algorithm translates to a different direction for each different choice of the norm.
More on the Gradient Vector

- All our ideas about first and second derivative in the case of a single variable carry over to the directional derivative.
- What does the gradient $\nabla f(x)$ tell you about the function $f(x)$? While there exist infinitely many direction vectors $v$ at any point $x$, there is a unique gradient vector $\nabla f(x)$.
- Since we expressed $D_v f(x)$ as the dot product of $\nabla f(x)$ with $v$, we can study $\nabla f(x)$ independently.

**Claim**

Suppose $f$ is a differentiable function of $x \in \mathbb{R}^n$. The maximum value of the directional derivative $D_v f(x)$ is $\|\nabla f(x)\|$ and it is so when $v$ has the same direction as the gradient vector $\nabla f(x)$. 
More on the Gradient Vector (contd.)

Proof:

The cauchy schwartz inequality when applied in the euclidian space gives us

\[ |\langle x, y \rangle| \leq \| x \| \| y \| \]

for any \( x, y \in \mathbb{R}^n \), with equality holding if \( x \) and \( y \) are in the same direction.
More on the Gradient Vector (contd.)

Proof:

- The **cauchy schwartz inequality** when applied in the euclidean space gives us
  $$|x \ y| \leq ||x|| \ ||y||$$ for any $$x, y \in \mathbb{R}$$, with equality holding if $$x$$ and $$y$$ are linearly dependent.

- The inequality gives upper and lower bounds on the dot product between two vectors;
  $$-||x|| \ ||y|| \leq x \ y \leq ||x|| \ ||y||$$.

- Applying these bounds to the right hand side of (??) and using the fact that $$||v|| = 1$$, we get
More on the Gradient Vector (contd.)

Proof:

- The *Cauchy-Schwarz inequality* when applied in the Euclidean space gives us $|\mathbf{x} \cdot \mathbf{y}| \leq ||\mathbf{x}|| ||\mathbf{y}||$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, with equality holding if $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent.

- The inequality gives upper and lower bounds on the dot product between two vectors: $-||\mathbf{x}|| ||\mathbf{y}|| \leq \mathbf{x} \cdot \mathbf{y} \leq ||\mathbf{x}|| ||\mathbf{y}||$.

- Applying these bounds to the right hand side of (??) and using the fact that $||\mathbf{v}|| = 1$, we get

$$-||\nabla f(\mathbf{x})|| \leq D_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v} \leq ||\nabla f(\mathbf{x})||$$

with equality holding if $\mathbf{v} = k \nabla f(\mathbf{x})$ for some $k \geq 0$.

- Since $||\mathbf{v}|| = 1$, equality can hold if $\mathbf{v} = \frac{\nabla f(\mathbf{x})}{||\nabla f(\mathbf{x})||}$.

This is L2 norm. H/w: How do you prove the other cases discussed in the class for other choices of norms.
Thus, the maximum rate of change of $f$ at a point $x$ is given by the norm $||\nabla f(x)||$ of the gradient vector at $x$.

And the direction in which the rate of change of $f$ is maximum is given by the unit vector $\frac{\nabla f(x)}{||\nabla f(x)||}$.

An associated fact is that the minimum value of the directional derivative $D_v f(x)$ is $-||\nabla f(x)||$ and it is attained when $v$ has the opposite direction of the gradient vector, i.e., $-\frac{\nabla f(x)}{||\nabla f(x)||}$.

The method of steepest descent uses this result to iteratively choose a new value of $x$ by traversing in the direction of $-\nabla f(x)$, especially while minimizing the value of some complex function.
Visualizing the Gradient Vector

Consider the function $f(x_1, x_2) = x e_1$. The Figure below shows 10 level curves for this function, corresponding to $f(x_1, x_2) = c$ for $c = 1, 2, \ldots, 10$.

The idea behind a level curve is that as you change $x$ along any level curve, the function value remains unchanged, but as you move $x$ across level curves, the function value changes.
Vanishing of the Directional Derivative

What if $D_v f(x)$ turns out to be 0?

Either gradient of $f$ is 0
OR
$v$ is orthogonal to the gradient

Gradient at $(1,1) = (2,2)$

Level curves for $x^2 + y^2$
Level Surface based Interpretation of Gradient: Examples

The level surfaces for $f(x, y, z) = x^2 + y^2 + z^2$ are shown in the Figure below. The gradient at $(1, 1, 1)$ is orthogonal to the tangent hyperplane to the level surface $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 = 3$ at $(1, 1, 1)$. The gradient vector at $(1, 1, 1)$ is $[2, 2, 2]^T$ and the tangent hyperplane has the equation $2(x - 1) + 2(x - 1) + 2(x - 1) = 0$, which is a plane in $3D$. 

![Level Surfaces Example](image)
Gradient and Convex Functions?

- How do we understand the behaviour of gradients for convex functions?
- While we have a lot to see in the coming sessions, here is a small peek through *sub-level sets* of a convex function

---

**Definition**

**[Sublevel Sets]:** Let \( D \subseteq \mathbb{R}^n \) be a nonempty set and \( f: D \rightarrow \mathbb{R} \). The set

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

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.
Gradient and Convex Functions?

- How do we understand the behaviour of gradients for convex functions?
- While we have a lot to see in the coming sessions, here is a small peek through sub-level sets of a convex function

**Definition**

**[Sublevel Sets]:** Let $D \subseteq \mathbb{R}^n$ be a nonempty set and $f : D \rightarrow \mathbb{R}$. The set

$$L_\alpha(f) = \{ x \mid x \in D, \ f(x) \leq \alpha \}$$

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

Now if a function $f$ is convex,

will the sublevel set be necessarily a convex set?