Convexity, Local and Global Optimality, etc.

Recap: Some Interesting Connections in $\Re^n$

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. $\mathcal{S}$ is closed if and only if $closure(\mathcal{S}) = \mathcal{S}$.
3. A bounded set can be defined in terms of a closed set; A set $\mathcal{S}$ 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 $\mathcal{S}$ is $closure(\mathcal{S}) = int(\mathcal{S}) \cup \partial(\mathcal{S})$.

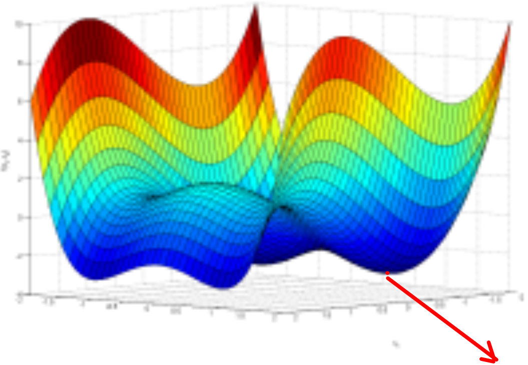
Extending Open, Closed sets, Boundary, Interior, etc to Topological Sets

This is for Optinal Reading

1. Recap: Open Set follows from Defintion 1 of Topology. Neighborhood follows from Definition 2 of Topology. By this definition, can point in interior be limit poi
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 atleast 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 iff $\forall\ \epsilon > 0$, $\{y \in S \text{ s.t. } 0 < d(y,x) < \epsilon\} \neq \emptyset\}$.
3. **Closure of S** $= closure(S) = S \cup \{\text{limit points of } S\}$.
4. **Boundary ∂S of S:** Is the subset of S such that every neighborhood of a point from ∂S contains atleast one point in S and one point not in S.
   - If S has a metric d then:
     $\partial S = \{x \in S | \forall\ \epsilon > 0, \exists\ y\ s.t.\ d(x,y) < \epsilon\ and\ y \in S\ and \exists\ z\ s.t.\ d(x,z) < \epsilon\ and\ z \notin S\}$
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 $f(x_1, x_2) = 3x_1^2 - x_1^3 - 2x_2^2 + x_2^4$. As can be seen in the plot, the function has several local maxima and minima.



A local min

Figure 1:

Convexity and Global Minimum

Fundamental chracteristics: **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, f(y) < f(x)

We are trying to prove by contradiction that a y different from x cannot exist

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 for all points in the epsilon disc, the value is >= f(x)

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 $\mathbf{z}$ 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 $f : \mathcal{D} \rightarrow \Re$ 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 $\mathbf{z} = \theta\mathbf{y} + (1 - \theta)\mathbf{x}$ 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

We have shown a specific value for theta when we assume a norm

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. □

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 should lie in D

Proof by contradiction

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.

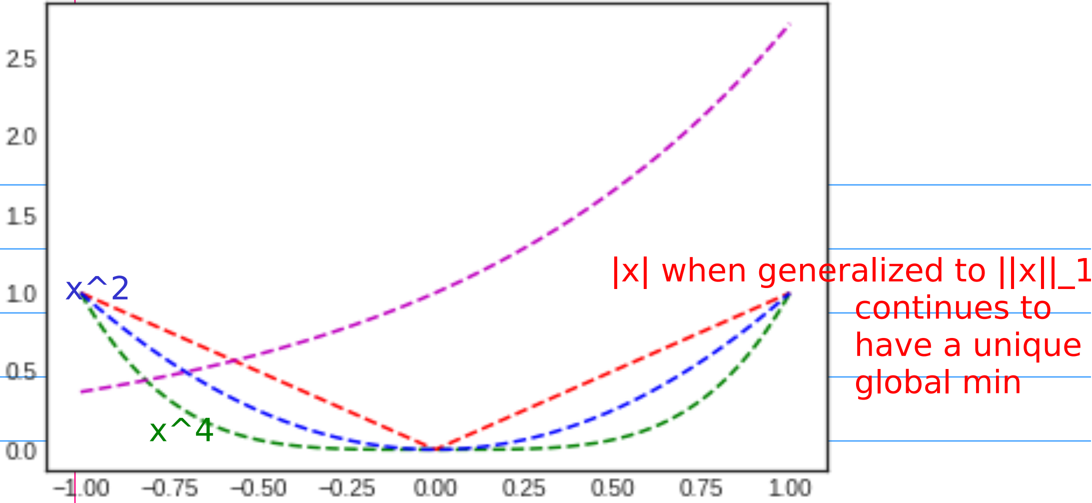
Theorem

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 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. □



It is possible that a convex function is NOT strictly convex and yet it has a unique global minimum

Convexity and Differentiability

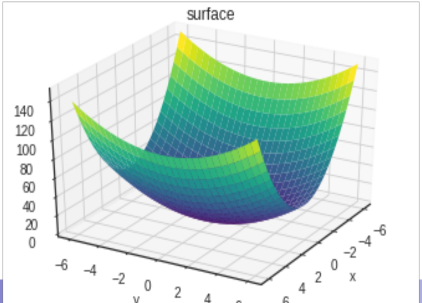
1. Recap for differentiable $f : \Re \rightarrow \Re$ the equivalent definition of convexity

A nondecreasing f'

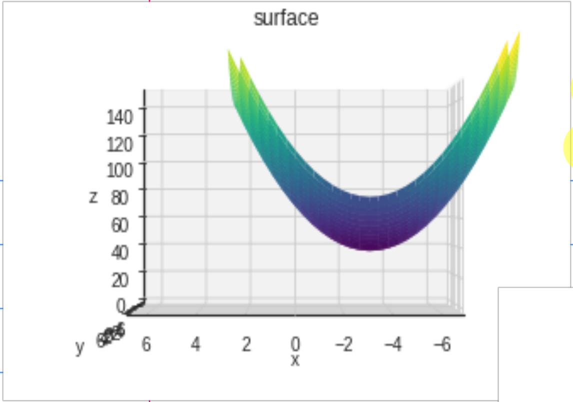
Convexity and Differentiability

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

3x^2 - x + y^2

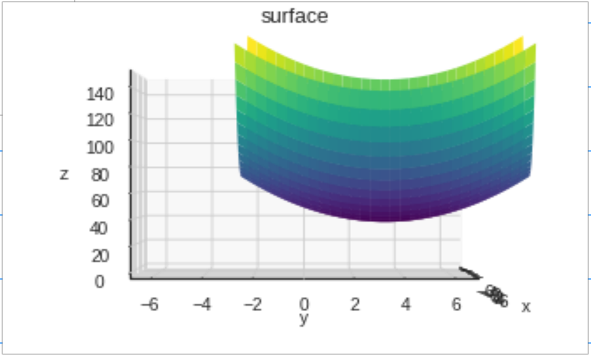


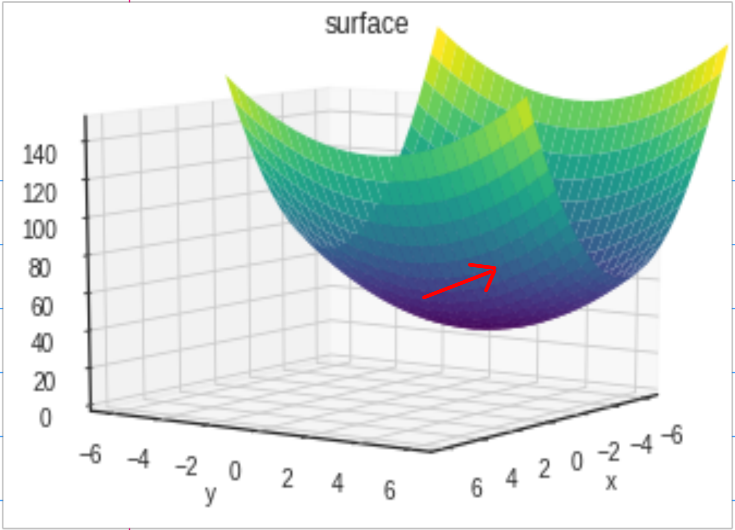
In both views, I find that the convexit
of the function is reflected in the
non-decreasing nature of the
derivatives along the respective axis
(directions)



View from x-axis

View from y-axis





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 $\Re^n$. These also include relationships between convex functions and convex sets, and first and second order conditions for convexity.

The Direction Vector

- Consider a function $f(\mathbf{x})$, 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 $f : \mathcal{D} \rightarrow \Re$, $\mathcal{D} \subseteq \Re^n$ 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

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

$$D_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} \tag{1}$$

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.

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

Claim

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

$$D_{\mathbf{v}} f(\mathbf{x}) = \sum_{k=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_k} v_k \tag{2}$$

Directional Derivative: Simplified Expression

Define $g(h) = f(\mathbf{x} + \mathbf{v}h)$. 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 $g(h) = f(\mathbf{x} + \mathbf{v}h)$. Now:

- $g'(0) = \lim_{h \to 0} \frac{g(0+h)-g(0)}{h} = \lim_{h \to 0} \frac{f(\mathbf{x}+h\mathbf{v})-f(\mathbf{x})}{h}$, which is the expression for the directional derivative defined in equation 1. Thus, $g'(0) = D_{\mathbf{v}} f(\mathbf{x})$.
- By definition of the chain rule for partial differentiation, we get another expression for $g'(0)$ as

$$g'(0) =$$

Directional Derivative: Simplified Expression

Define $g(h) = f(\mathbf{x} + \mathbf{v}h)$. Now:

- $g'(0) = \lim_{h\to 0} \frac{g(0+h)-g(0)}{h} = \lim_{h\to 0} \frac{f(\mathbf{x}+h\mathbf{v})-f(\mathbf{x})}{h}$, which is the expression for the directional derivative defined in equation 1. Thus, $g'(0) = D_{\mathbf{v}} f(\mathbf{x})$.
- By definition of the chain rule for partial differentiation, we get another expression for $g'(0)$ as

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

Therefore, $g'(0) = D_{\mathbf{v}} f(\mathbf{x}) = \sum_{k=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_k} v_k$ □

Homeworks:

1. Consider the polynomial $f(x,y,z) = x^2y + z\sin xy$ and the unit vector $\mathbf{v}^T = \frac{1}{\sqrt{3}}[1,1,1]^T$. Consider the point $p_0 = (0,1,3)$. Compute the directional derivative of f at p_0 in the direction of $\mathbf{v}$.
2. Compute the rate of change of $f(x,y,z) = e^{xyz}$ at $p_0 = (1,2,3)$ in the direction from $p_1 = (1,2,3)$ to $p_2 = (-4,6,-1)$.

Illustrating Computation of Directional Derivative

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

Illustrating Computation of Directional Derivative

- Consider the polynomial $f(x, y, z) = x^2 y + z \sin xy$ and the unit vector $\mathbf{v}^T = \frac{1}{\sqrt{3}}[1, 1, 1]^T$. Consider the point $p_0 = (0, 1, 3)$. We will compute the directional derivative of f at p_0 in the direction of $\mathbf{v}$.
- To do this, we first compute the gradient of f in general:
  $\nabla f = \left[2xy + yz \cos xy,\ x^2 + xz \cos xy,\ \sin xy\right]^T$.

Illustrating Computation of Directional Derivative

- Consider the polynomial $f(x, y, z) = x^2 y + z \sin xy$ and the unit vector $\mathbf{v}^T = \frac{1}{\sqrt{3}}[1, 1, 1]^T$. Consider the point $p_0 = (0, 1, 3)$. We will compute the directional derivative of f at p_0 in the direction of $\mathbf{v}$.
- To do this, we first compute the gradient of f in general: $\nabla f = \left[2xy + yz \cos xy,\ x^2 + xz \cos xy,\ \sin xy\right]^T$.
- Evaluating the gradient at a specific point p_0, $\nabla f(0, 1, 3) = [3,\ 0,\ 0]^T$. The directional derivative at p_0 in the direction $\mathbf{v}$ is $D_{\mathbf{v}} f(0, 1, 3) = [3, 0, 0] . \frac{1}{\sqrt{3}}[1, 1, 1]^T = \sqrt{3}$.
- This directional derivative is the rate of change of f at p_0 in the direction $\mathbf{v}$; it is positive indicating that the function f increases at p_0 in the direction $\mathbf{v}$.

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 $\nabla f(\mathbf{x})$ tell you about the function $f(\mathbf{x})$? While there exist infinitely many direction vectors $\mathbf{v}$ at any point $\mathbf{x}$, there is a unique gradient vector $\nabla f(\mathbf{x})$.
- Since we expressed $D_{\mathbf{v}} f(\mathbf{x})$ as the dot product of $\nabla f(\mathbf{x})$ with $\mathbf{v}$, we can study $\nabla f(\mathbf{x})$ 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 $\nabla f(\mathbf{x})$ tell you about the function $f(\mathbf{x})$? While there exist infinitely many direction vectors $\mathbf{v}$ at any point $\mathbf{x}$, there is a unique gradient vector $\nabla f(\mathbf{x})$.
- Since we expressed $D_{\mathbf{v}} f(\mathbf{x})$ as the dot product of $\nabla f(\mathbf{x})$ with $\mathbf{v}$, we can study $\nabla f(\mathbf{x})$ independently.

Claim

Suppose f is a differentiable function of $\mathbf{x} \in \Re^n$. The maximum value of the directional derivative $D_{\mathbf{v}} f(\mathbf{x})$ is ||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 $\nabla f(\mathbf{x})$ tell you about the function $f(\mathbf{x})$? While there exist infinitely many direction vectors $\mathbf{v}$ at any point $\mathbf{x}$, there is a unique gradient vector $\nabla f(\mathbf{x})$.
- Since we expressed $D_{\mathbf{v}} f(\mathbf{x})$ as the dot product of $\nabla f(\mathbf{x})$ with $\mathbf{v}$, we can study $\nabla f(\mathbf{x})$ independently.

Claim

Suppose f is a differentiable function of $\mathbf{x} \in \Re^n$. The maximum value of the directional derivative $D_{\mathbf{v}} f(\mathbf{x})$ is $||\nabla f(\mathbf{x}||$ and it is so when $\mathbf{v}$ has the same direction as the gradient vector $\nabla f(\mathbf{x})$.

More on the Gradient Vector (contd.)

Proof:

- The *cauchy schwartz inequality* when applied in the eucledian space gives us $|\mathbf{x}^T\mathbf{y}| \leq ||\mathbf{x}||||\mathbf{y}||$ for any $\mathbf{x}, \mathbf{y} \in \Re^n$, with equality holding *iff* x and y are in the same direction

More on the Gradient Vector (contd.)

Proof:

- The *cauchy schwartz inequality* when applied in the eucledian space gives us $|\mathbf{x}^T\mathbf{y}| \leq ||\mathbf{x}||||\mathbf{y}||$ for any $\mathbf{x}, \mathbf{y} \in \Re^n$, with equality holding *iff* $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent.
- The inequality gives upper and lower bounds on the dot product between two vectors; $-||\mathbf{x}||||\mathbf{y}|| \leq \mathbf{x}^T\mathbf{y} \leq ||\mathbf{x}||||\mathbf{y}||$.
- Applying these bounds to the right hand side of (**??**) and using the fact that $||\mathbf{v}|| = 1$, we get

More on the Gradient Vector (contd.)

Proof:

- The *cauchy schwartz inequality* when applied in the eucledian space gives us $|\mathbf{x}^T\mathbf{y}| \leq ||\mathbf{x}||||\mathbf{y}||$ for any $\mathbf{x}, \mathbf{y} \in \Re^n$, with equality holding *iff* $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent.
- The inequality gives upper and lower bounds on the dot product between two vectors; $-||\mathbf{x}||||\mathbf{y}|| \leq \mathbf{x}^T\mathbf{y} \leq ||\mathbf{x}||||\mathbf{y}||$.
- Applying these bounds to the right hand side of (**??**) and using the fact that $||\mathbf{v}|| = 1$, we get

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

  with equality holding *iff* $\mathbf{v} = k\nabla f(\mathbf{x})$ for some $k \geq 0$.
- Since $||\mathbf{v}|| = 1,$ equality can hold *iff* $\mathbf{v} = \frac{\nabla f(\mathbf{x})}{||\nabla f(\mathbf{x})||}$.

This is L2 norm. H/w: How do you prove the other cases discussed in the class for other choices of norms

More on the Gradient Vector (contd.)

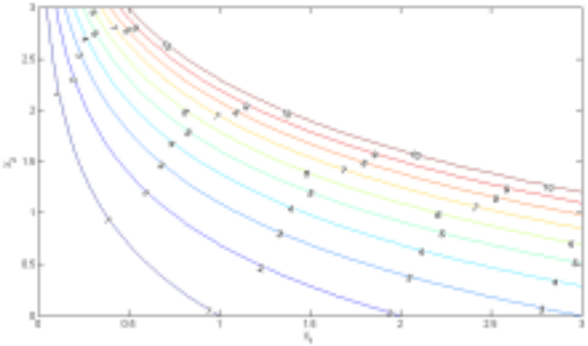
- Thus, the maximum rate of change of f at a point $\mathbf{x}$ is given by the norm $||\nabla f(\mathbf{x}||$ of the gradient vector at $\mathbf{x}$.
- And the direction in which the rate of change of f is maximum is given by the unit vector $\frac{\nabla f(\mathbf{x}}{||\nabla f(\mathbf{x}||}$.
- An associated fact is that the minimum value of the directional derivative $D_{\mathbf{v}} f(\mathbf{x})$ is $-||\nabla f(\mathbf{x})||$ and it is attained when $\mathbf{v}$ has the opposite direction of the gradient vector, *i.e.*, $-\frac{\nabla f(\mathbf{x}}{||\nabla f(\mathbf{x}||}$.

using L2 norm

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

Visualizing the Gradient Vector

Consider the function $f(x_1, x_2) = x_1 e^{x_2}$. 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 $\mathbf{x}$ along any level curve, the function value remains unchanged, but as you move $\mathbf{x}$ across level curves, the function value changes.

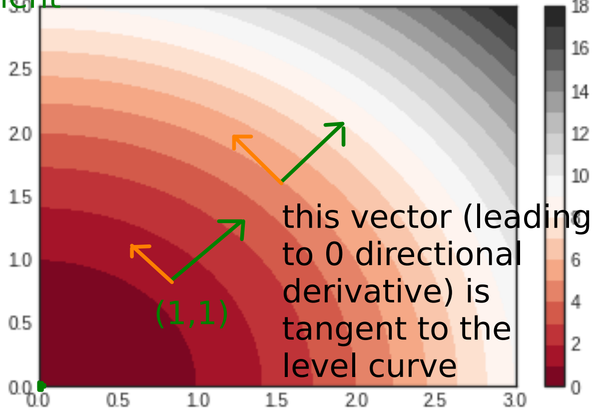
Vanishing of the Directional Derivative

What if $D_{\mathbf{v}} f(\mathbf{x})$ 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



Vanishing of the Directional Derivative

What if $D_{\mathbf{v}} f(\mathbf{x})$ turns out to be 0?
We then expect that $\nabla f(\mathbf{x})$ and $\mathbf{v}$ are othogonal.

Definition

Level Surface/Set: The *level surface/set* of $f(\mathbf{x})$ at $\mathbf{x}^*$ is

$$\{\mathbf{x} | f(\mathbf{x}) = f(\mathbf{x}^*)\} \tag{3}$$

Vanishing of the Directional Derivative

What if $D_{\mathbf{v}} f(\mathbf{x})$ turns out to be 0?
We then expect that $\nabla f(\mathbf{x})$ and $\mathbf{v}$ are othogonal.

Definition

Level Surface/Set: The *level surface/set* of $f(\mathbf{x})$ at $\mathbf{x}^*$ is

$$\{\mathbf{x} | f(\mathbf{x}) = f(\mathbf{x}^*)\} \tag{3}$$

There is a useful result in this regard.

Claim

Let $f : \mathcal{D} \rightarrow \Re$ with $\mathcal{D} \in \Re^n$ be a differentiable function. The gradient ∇f evaluated at $\mathbf{x}^*$ is orthogonal to the tangent hyperplane (tangent line in case $n = 2$) to the level surface of f passing through $\mathbf{x}^*$.

Vanishing of the Directional Derivative & Level Surfaces: Proof

Proof: Let $\mathcal{K}$ be the range of f and let $k \in \mathcal{K}$ such that $f(\mathbf{x}^*) = k$.

- Consider the level surface $f(\mathbf{x}) = k$. Let $\mathbf{r}(t) = [x_1(t), x_2(t), \ldots, x_n(t)]$ be a curve on the level surface, parametrized by $t \in \Re$, with $\mathbf{r}(0) = \mathbf{x}^*$.
- Then, $f(x(t), y(t), z(t)) = k$. Applying the chain rule

$$\frac{df(\mathbf{r}(t))}{dt} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{dx_i(t)}{dt} = \nabla^T f(\mathbf{x}(t))\frac{d\mathbf{r}(t)}{dt} = 0$$

- For $t = 0$, the equations become

$$\nabla^T f(\mathbf{x}^*)\frac{d\mathbf{r}(0)}{dt} = 0$$

- Now, $\frac{d\mathbf{r}(t)}{dt}$ represents any tangent vector to the curve through $\mathbf{r}(t)$ which lies completely on the level surface.

Vanishing of the Directional Derivative & Level Surfaces: Proof

$$\nabla^T f(\mathbf{x}^*) \frac{d\mathbf{r}(0)}{dt} = 0$$

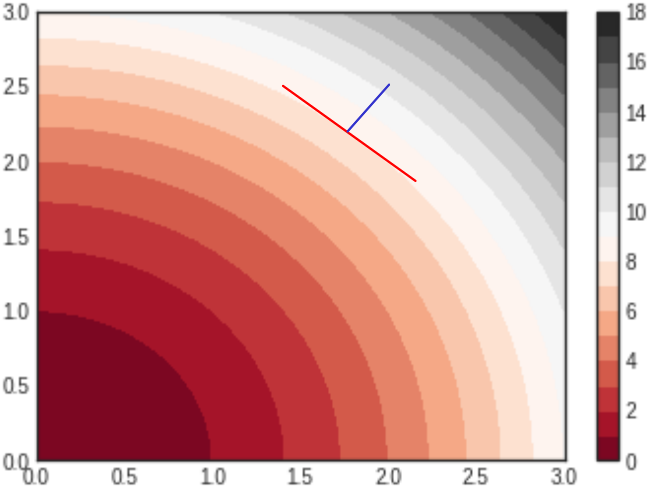
- That is, the tangent line to any curve at $\mathbf{x}^*$ on the level surface containing $\mathbf{x}^*$, is orthogonal to $\nabla f(\mathbf{x}^*)$.
- Since the tangent hyperplane to a surface at any point is the hyperplane containing all tangent vectors to curves on the surface passing through the point, the gradient $\nabla f(\mathbf{x}^*)$ is perpendicular to the tangent hyperplane to the level surface passing through that point $\mathbf{x}^*$.
- The equation of the tangent hyperplane is given by

Vanishing of the Directional Derivative & Level Surfaces: Proof

$$\nabla^T f(\mathbf{x}^*) \frac{d\mathbf{r}(0)}{dt} = 0$$

- That is, the tangent line to any curve at $\mathbf{x}^*$ on the level surface containing $\mathbf{x}^*$, is orthogonal to $\nabla f(\mathbf{x}^*)$.
- Since the tangent hyperplane to a surface at any point is the hyperplane containing all tangent vectors to curves on the surface passing through the point, the gradient $\nabla f(\mathbf{x}^*)$ is perpendicular to the tangent hyperplane to the level surface passing through that point $\mathbf{x}^*$.
- The equation of the tangent hyperplane is given by $(\mathbf{x} - \mathbf{x}^*)^T \nabla f(\mathbf{x}^*) = 0$

This dot product will appear in definition of convexity, quasi-convexity, ...



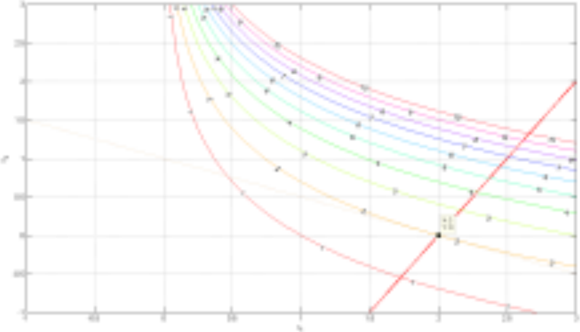
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
18
16
14
12
10
8
6
4
2
0

Level Surface based Interpretation of Gradient

- Recall that the normal to a plane can be found by taking the cross product of any two vectors lying within the plane. Thus, the gradient vector $\nabla f(\mathbf{x}^*)$ at any point $\mathbf{x}^*$ on the level surface of a function $f(.)$ is **normal to the tangent hyperplane (or tangent line in the case of two variables) to the surface at the same point.**
- The same gradient vector $\nabla f(\mathbf{x}^*)$ at a point $\mathbf{x}^*$ can also be conveniently computed as the **vector of partial derivatives of the function at that point.**
- We will illustrate this geometric understanding through some examples.

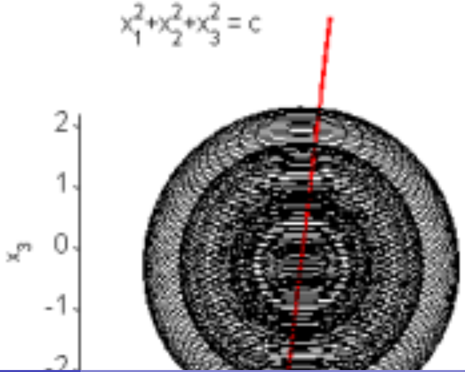
Level Surface based Interpretation of Gradient: Examples

- Consider the same plot as earlier with a gradient vector at $(2,0)$ as shown below. The gradient vector $[1,\ 2]^T$ is perpendicular to the tangent hyperplane to the level curve $x_1 e^{x_2} = 2$ at $(2,0)$. The equation of the tangent hyperplane is $(x_1 - 2) + 2(x_2 - 0) = 0$ and it turns out to be a tangent line.



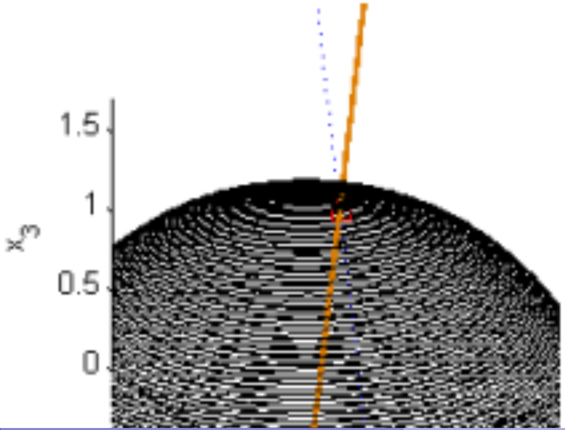
Level Surface based Interpretation of Gradient: Examples

The level surfaces for $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ 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 tanget hyperplane has the equation $2(x_1 - 1) + 2(x_2 - 1) + 2(x_3 - 1) = 0$, which is a plane in $3D$.



Level Surface based Interpretation of Gradient: Examples

On the other hand, the dotted line in the Figure below is not orthogonal to the level surface, since it does not coincide with the gradient.



Level Surface based Interpretation of Gradient: Examples

Determine the equations of
(a) the tangent plane to the paraboloid $\mathcal{P} : x_1 = x_2^2 + x_3^2 + 2$ at $(-1, 1, 0)$ and
(b) the normal line to the tangent plane.

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 $\mathcal{D} \subseteq \Re^n$ 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,

will the sublevel set be necessarily 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 $\mathcal{D} \subseteq \Re^n$ be a nonempty set and $f : \mathcal{D} \rightarrow \Re$. The set

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

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. □

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 $\left\{(x_1, x_2) \mid x_2 \leq 0\right\}$, which is convex. However, the function $f(\mathbf{x})$ itself is not convex.