Convexity, Local and Global Optimality, etc.

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

Homeworks:

1. 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 $p_0 = (0, 1, 3)$. Compute the directional derivative of f at p_0 in the direction of $\mathbf{v}$.
2. find 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 $p_2 = (-4, 6, -1)$.

The Gradient Vector and Directional Derivative

- We can see that the right hand side of (2) can be realized as the dot product of two vectors, *viz.*, $\left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n}\right]^T$ and $\mathbf{v}$.
- Let us denote $\frac{\partial f(\mathbf{x})}{\partial x_i}$ by $f_{x_i}(\mathbf{x})$. Then we assign a name to this special vector:

Definition

[Gradient Vector]: If f is differentiable function of $\mathbf{x} \in \Re^n$, then the gradient of $f(\mathbf{x})$ is the vector function $\nabla f(\mathbf{x})$, defined as:

$$\nabla f(\mathbf{x}) = \left[f_{x_1}(\mathbf{x}),\ f_{x_2}(\mathbf{x}), \ldots,\ f_{x_n}(\mathbf{x})\right]$$

The directional derivative of a function f at a point $\mathbf{x}$ in the direction of a unit vector $\mathbf{v}$ can be now written as

The Gradient Vector and Directional Derivative

- We can see that the right hand side of (2) can be realized as the dot product of two vectors, *viz.*, $\left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n}\right]^T$ and $\mathbf{v}$.
- Let us denote $\frac{\partial f(\mathbf{x})}{\partial x_i}$ by $f_{x_i}(\mathbf{x})$. Then we assign a name to this special vector:

Definition

[Gradient Vector]: If f is differentiable function of $\mathbf{x} \in \Re^n$, then the gradient of $f(\mathbf{x})$ is the vector function $\nabla f(\mathbf{x})$, defined as:

$$\nabla f(\mathbf{x}) = \left[f_{x_1}(\mathbf{x}),\ f_{x_2}(\mathbf{x}), \ldots,\ f_{x_n}(\mathbf{x})\right]$$

The directional derivative of a function f at a point $\mathbf{x}$ in the direction of a unit vector $\mathbf{v}$ can be now written as

$$D_{\mathbf{v}} f(\mathbf{x}) = \nabla^T f(\mathbf{x}).\mathbf{v} \tag{3}$$

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 $\mathbf{v}$.

Illustrating Computation of Directional Derivative

- As another example, let us find 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

- As another example, let us find 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 $p_2 = (-4, 6, -1)$.
- We first construct a unit vector from p_1 to p_2; $\mathbf{v} = \frac{1}{\sqrt{57}}[-5, 4, -4]$.
- The gradient of f in general is $\nabla f = [yze^{xyz},\ xze^{xyz},\ xye^{xyz}] = e^{xyz}[yz,\ xz,\ xy]$.
- Evaluating the gradient at a specific point p_0, $\nabla f(1, 2, 3) = e^6\,[6, 3, 2]^T$. The directional derivative at p_0 in the direction $\mathbf{v}$ is $D_{\mathbf{u}} f(1, 2, 3) = e^6[6, 3, 2].\frac{1}{\sqrt{57}}[-5, 4, -4]^T = e^6 \frac{-26}{\sqrt{57}}$.
- This directional derivative is negative, indicating that the function f decreases at p_0 in the direction from p_1 to p_2.

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.

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* $\nabla f(\mathbf{x})$.

Proof: Directional derivative is a dot product of gradient and directior
which can be upper bounded using Cauchy Shwarz inequality

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 (3) and using the fact that $||\mathbf{v}|| = 1$, we get

Upper and lower bounds for directional derivative at x.

Are these upper and lower bounds attainable?

ANS: Yes. In or against direction of gradient of f at 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* $\mathbf{x}$ 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 (3) and using the fact that $||\mathbf{v}|| = 1$, we get

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

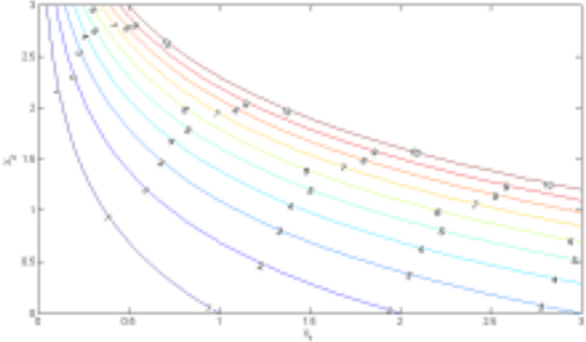
  with equality holding *iff* $\mathbf{v} = k\nabla f(\mathbf{x})$ for some $k \geq 0$.
- Since $||\mathbf{v}|| = 1$, equality can hold *iff* $\mathbf{v} = \frac{\nabla f(\mathbf{x})}{||\nabla f(\mathbf{x})||}$.

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.*, $-\frac{\nabla f(\mathbf{x}}{||\nabla f(\mathbf{x}||}$.
- 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

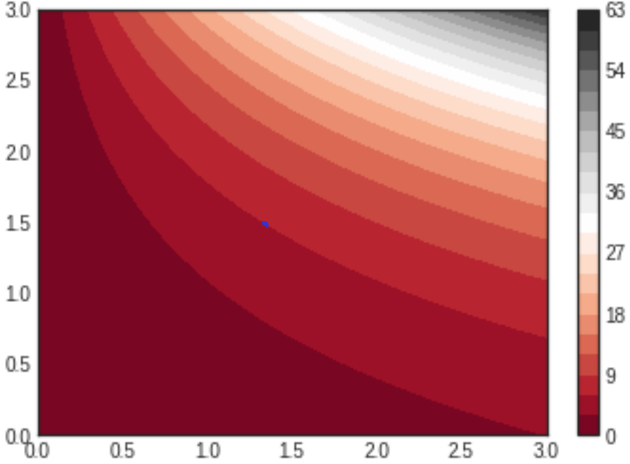
Consider the function $f(x_1, x_2) = x_1 e^{x_2}$. The Figure below shows 10 level curves for this function, corresponding to $f(x_1, x_2) = c$ for $c = 1, 2, \ldots, 10$.



will be useful and discussed for constrained optimization

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.

will be discussed now (for function being minimized

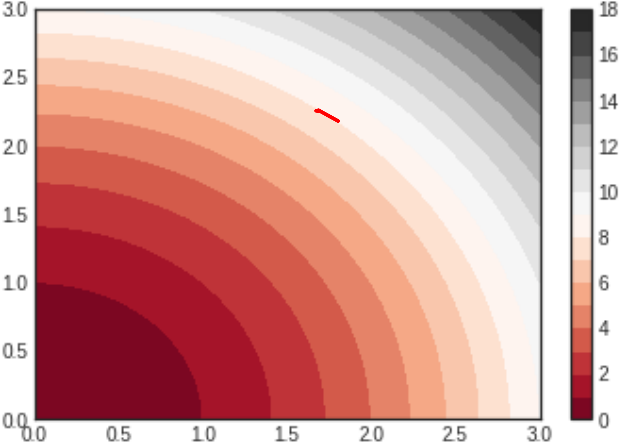


Level curves for
f(x,y) = x * exp(y)

Vanishing of the Directional Derivative

What if $D_{\mathbf{v}} f(\mathbf{x})$ turns out to be 0?

Level curves for x^2 + y^2



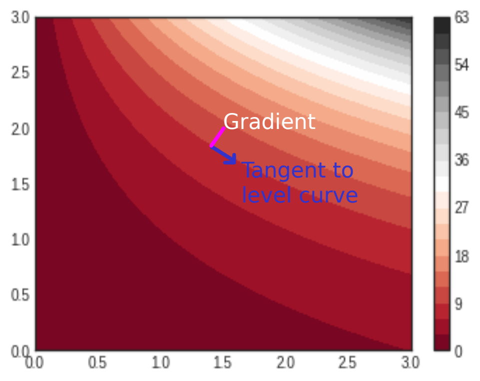
Expect directional derivative to be 0 in direction tangential to any level curve

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



$$\{\mathbf{x} | f(\mathbf{x}) = f(\mathbf{x}^*)\} \tag{4}$$

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

$$\{\mathbf{x} | f(\mathbf{x}) = f(\mathbf{x}^*)\} \tag{4}$$

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 $\mathbf{x}^*$.

Intuition... We consider any parametric curve passing through x*
and lying on the level set.. has gradient orthogonal to tangent line

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

For this example in 3d, df/dt = dot product of gradient of f with the vector of derivatives of xi wrt t

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

Vanishing of the Directional Derivative & Level Surfaces: Proof (Not rigorous)

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.

The tangent plane is the plane containing all such tangent vectors across all such r(t)

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

Hyperplane(x*,gradient of f at x*)

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 $(\mathbf{x} - \mathbf{x}^*)^T \nabla f(\mathbf{x}^*) = 0$

Vanishing of the Directional Derivative & Level Surfaces: Proof

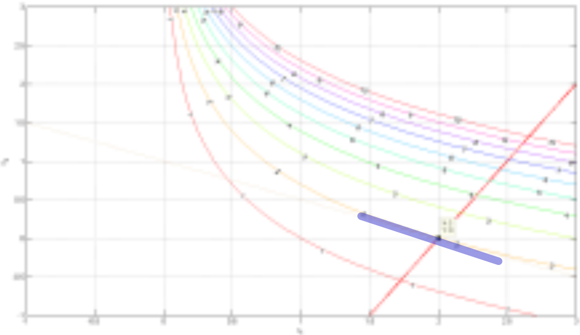
- ..

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.

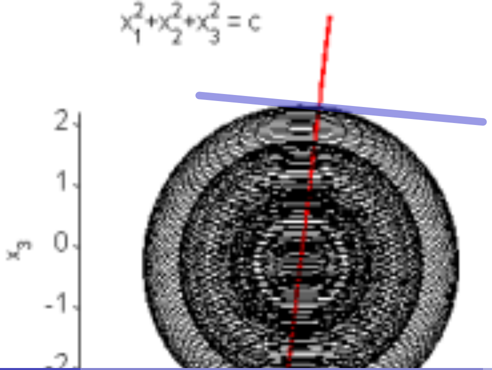
Level Surface based Interpretation of Gradient: Examples

- 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.



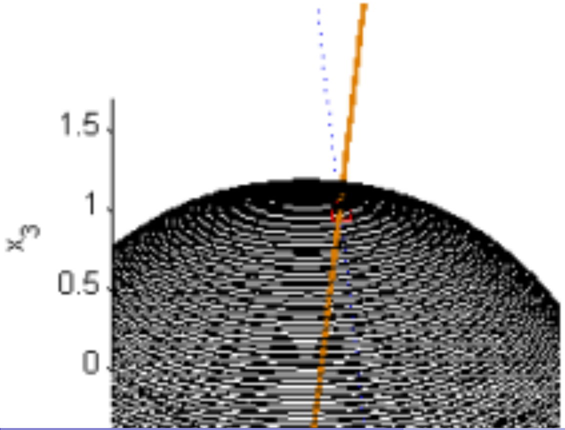
Level Surface based Interpretation of Gradient: Examples

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

- Let $f(x_1, x, x_3) = x_1^2 x_2^3 x_3^4$ and consider the point $\mathbf{x}^0 = (1, 2, 1)$. We will find the equation of the tangent plane to the level surface through $\mathbf{x}^0$.

Compute gradient at x0

Level Surface based Interpretation of Gradient: Examples

- Let $f(x_1, x, x_3) = x_1^2 x_2^3 x_3^4$ and consider the point $\mathbf{x}^0 = (1, 2, 1)$. We will find the equation of the tangent plane to the level surface through $\mathbf{x}^0$.
- The level surface through $\mathbf{x}^0$ is determined by setting f equal to its value evaluated at $\mathbf{x}^0$; that is, the level surface will have the equation $x_1^2 x_2^3 x_3^4 = 1^2 2^3 1^4 = 8$.
- The gradient vector (normal to tangent plane) at $(1, 2, 1)$ is $\nabla f(x_1, x_2, x_3)\big|_{(1,2,1)} = [2x_1 x_2^3 x_3^4,\ 3x_1^2 x_2^2 x_3^4, 4x_1^2 x_2^3 x_3^3]^T\Big|_{(1,2,1)} = [16,\ 12,\ 32]^T$.
- The equation of the tangent plane at $\mathbf{x}^0$, given the normal vector $\nabla f(\mathbf{x}^0)$ can be easily written down: $\nabla f(\mathbf{x}^0)^T.[\mathbf{x} - \mathbf{x}^0] = 0$ which turns out to be $16(x_1 - 1) + 12(x_2 - 2) + 32(x_3 - 1) = 0$, a plane in $3D$.

Level Surface based Interpretation of Gradient: Examples

- Consider the function $f(x, y, z) = \frac{x}{y+z}$. The directional derivative of f in the direction of the vector $\mathbf{v} = \frac{1}{\sqrt{14}}[1,\ 2,\ 3]$ at the point $x^0 = (4, 1, 1)$ is

Level Surface based Interpretation of Gradient: Examples

- Consider the function $f(x, y, z) = \frac{x}{y+z}$. The directional derivative of f in the direction of the vector $\mathbf{v} = \frac{1}{\sqrt{14}}[1,\ 2,\ 3]$ at the point $x^0 = (4, 1, 1)$ is $\nabla^T f\Big|_{(4,1,1)} . \frac{1}{\sqrt{14}}[1,\ 2,\ 3]^T = \left[\frac{1}{y+z},\ -\frac{x}{(y+z)^2},\ -\frac{x}{(y+z)^2}\right]\Bigg|_{(4,1,1)} . \frac{1}{\sqrt{14}}[1,\ 2,\ 3]^T = \left[\frac{1}{2},\ -1,\ -1\right] . \frac{1}{\sqrt{14}}[1,\ 2,\ 3]^T = -\frac{9}{2\sqrt{14}}$.
- The directional derivative is negative, indicating that the function decreases along the direction of $\mathbf{v}$. Based on an earlier result, we know that the maximum rate of change of a function at a point $\mathbf{x}$ is given by $||\nabla f(\mathbf{x})||$ and it is in the direction $\frac{\nabla f(\mathbf{x})}{||\nabla f(\mathbf{x})||}$.
- In the example under consideration, this maximum rate of change at $\mathbf{x}^0$ is $\frac{3}{2}$ and it is in the direction of the vector $\frac{2}{3}\left[\frac{1}{2},\ -1,\ -1\right]$.

Level Surface based Interpretation of Gradient: Examples

Let us find the maximum rate of change of the function $f(x, y, z) = x^2y^3z^4$ at the point $\mathbf{x}^0 = (1, 1, 1)$ and the direction in which it occurs. The gradient at $\mathbf{x}^0$ is

Level Surface based Interpretation of Gradient: Examples

Let us find the maximum rate of change of the function $f(x, y, z) = x^2 y^3 z^4$ at the point $\mathbf{x}^0 = (1, 1, 1)$ and the direction in which it occurs. The gradient at $\mathbf{x}^0$ is $\nabla^T f\Big|_{(1,1,1)} = [2,\ 3,\ 4]$. The maximum rate of change at $\mathbf{x}^0$ is therefore $\sqrt{29}$ and the direction of the corresponding rate of change is $\frac{1}{\sqrt{29}}[2,\ 3,\ 4]$. The minimum rate of change is $-\sqrt{29}$ and the corresponding direction is $-\frac{1}{\sqrt{29}}[2,\ 3,\ 4]$.

Level Surface based Interpretation of Gradient: Examples

Let us 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.
To realize this as the level surface of a function of three variables, we define the function

f (such that P corresponds to one of its level sets)

Level Surface based Interpretation of Gradient: Examples

Let us 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.
To realize this as the level surface of a function of three variables, we define the function $f(x_1, x_2, x_3) = x_1 - x_2^2 - x_3^2$ and find that the paraboloid $\mathcal{P}$ is the same as the level surface $f(x_1, x_2, x_3) = +2$. The normal to the tangent plane to $\mathcal{P}$ at $\mathbf{x}^0$ is in the direction of the gradient vector $\nabla f(\mathbf{x}^0) = [1, -2, 0]^T$ and its parametric equation is
$[x_1,\ x_2,\ x_3] = [-1 + t,\ 1 - 2t,\ 0]$.
The equation of the tangent plane is therefore $(x_1 + 1) - 2(x_2 - 1) = 0$.

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

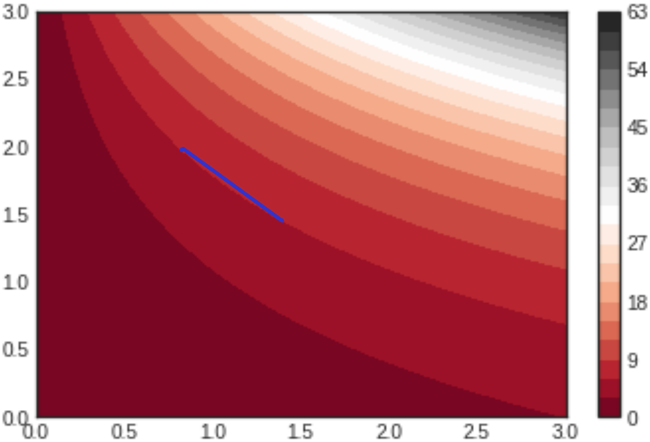
Definition

[Sublevel Sets]: Let $\mathcal{D} \subseteq \Re^n$ be a nonempty set and $f : \mathcal{D} \to \Re$. The set

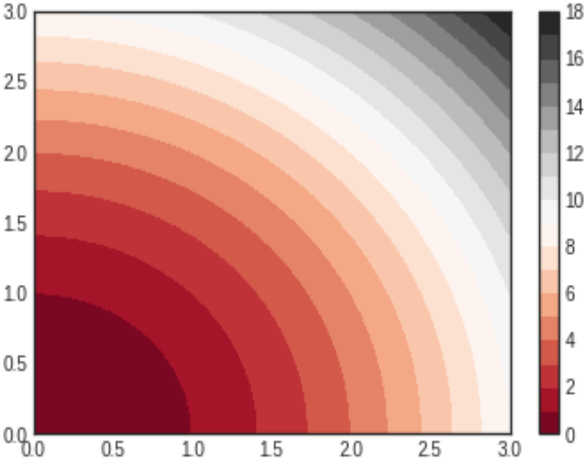
$$L_\alpha(f) = \left\{\mathbf{x} | \mathbf{x} \in \mathcal{D},\ f(\mathbf{x}) \leq \alpha\right\}$$

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

Now if a function f is convex, each of its sublevel set will be convex



Level sets for x1*exp(x2)
Sublevel sets are not convex
==> function cannot be
convex



x1^2 + x2^2
is a convex function
and so are its sublevel sets

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

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

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

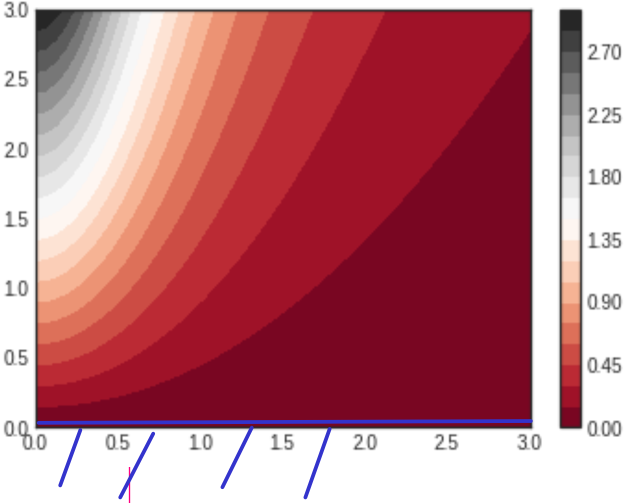
Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \rightarrow \Re$ be a convex function. Then $L_\alpha(f)$ is a convex set for any $\alpha \in \Re$.

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



0 level subset which is convex

Convex Function ⇒ 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.

A function is called quasi-convex if all its sub-level sets are convex sets Eg: Negative of the normal distribution $-\frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ is quasi-convex but not convex. [Homework]

Gradient, Convex Functions and Sub-level sets: A First Peek

We have already seen that

- The gradient $\nabla f(\mathbf{x}^*)$ at $\mathbf{x}^*$ is normal to the tangent hyperplane to the level set $\{\mathbf{x}|f(\mathbf{x}) = f(\mathbf{x}^*)\}$ at $\mathbf{x}^*$
- The gradient $\nabla f(\mathbf{x}^*)$ at $\mathbf{x}^*$ points in direction of increasing values of $f(.)$ at $\mathbf{x}^*$

Now, if $f(\mathbf{x})$ is also convex

Gradient, Convex Functions and Sub-level sets: A First Peek

We have already seen that

- The gradient $\nabla f(\mathbf{x}^*)$ at $\mathbf{x}^*$ is normal to the tangent hyperplane to the level set $\{\mathbf{x}|f(\mathbf{x}) = f(\mathbf{x}^*)\}$ at $\mathbf{x}^*$
- The gradient $\nabla f(\mathbf{x}^*)$ at $\mathbf{x}^*$ points in direction of increasing values of $f(.)$ at $\mathbf{x}^*$

Now, if $f(\mathbf{x})$ is also convex

- The gradient $\nabla f(\mathbf{x}^*)$ **at $\mathbf{x}^*$ is normal to the tangent hyperplane to the sub-level set $\{\mathbf{x}|f(\mathbf{x}) \leq f(\mathbf{x}^*)\}$ at $\mathbf{x}^*$**
- The tangent hyperplane defined by $\nabla f(\mathbf{x}^*)$ at $\mathbf{x}^*$ is a **supporting hyperplane** to the convex set $\{\mathbf{x}|f(\mathbf{x}) \leq f(\mathbf{x}^*)\}$ at $\mathbf{x}^*$