Functions of Several Variables

Hayk Aprikyan, Hayk Tarkhanyan

Functions of Several Variables

At the moment, we know how to be an effective apple salesperson. What about selling more than one item:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

i.e. how to deal with functions with more than one variable?

Functions of Several Variables

At the moment, we know how to be an effective apple salesperson. What about selling more than one item:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

i.e. how to deal with functions with more than one variable?

First, our input is now not a single number, but a pair of numbers (x, y).

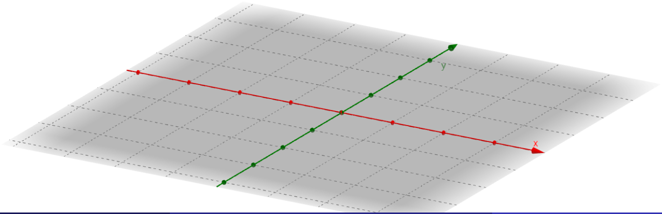
Functions of Several Variables

At the moment, we know how to be an effective apple salesperson. What about selling more than one item:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

i.e. how to deal with functions with more than one variable?

First, our input is now not a single number, but a pair of numbers (x, y). To visualize them, we need xy-plane



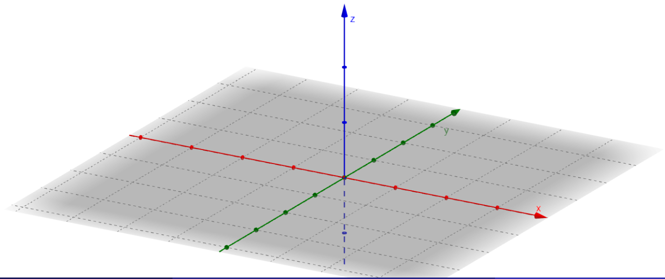
Functions of Several Variables

At the moment, we know how to be an effective apple salesperson. What about selling more than one item:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

i.e. how to deal with functions with more than one variable?

First, our input is now not a single number, but a pair of numbers (x, y). To visualize them, we need xy-plane + a new axis



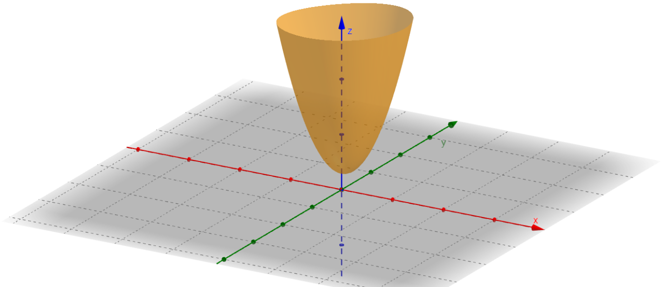
Functions of Several Variables

At the moment, we know how to be an effective apple salesperson. What about selling more than one item:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

i.e. how to deal with functions with more than one variable?

First, our input is now not a single number, but a pair of numbers (x, y). To visualize them, we need xy-plane + a new axis (the graph is a surface):



Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two
  - No idea of increasing/decreasing

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two
  - No idea of increasing/decreasing
- **Similarities:**

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two
  - No idea of increasing/decreasing
- **Similarities:**
  - Anything else is pretty much the same.

Functions of Several Variables

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two
  - No idea of increasing/decreasing
- **Similarities:**
  - Anything else is pretty much the same.

A picture is worth a thousand words

A typical graph in this case looks much like a napkin hanging in the air.

Note that compared to the single-variable case, now we have:

- **Differences:**
  - Two axes instead of one
  - Infinitely many directions instead of two
  - No idea of increasing/decreasing
- **Similarities:**
  - Anything else is pretty much the same.

A picture is worth a thousand words
(two pictures = two thousand words)

Partial Derivative

Again, suppose x and y are the costs of apples and peaches, and your profit is given by:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

Question

How can you measure the effect of increasing the cost of apple by a little (i.e. how quickly will f change if x changes)?

Partial Derivative

Again, suppose x and y are the costs of apples and peaches, and your profit is given by:

$$f(x, y) = x^2 + y^2 + e^{y-x}$$

Question

How can you measure the effect of increasing the cost of apple by a little (i.e. how quickly will f change if x changes)?

By fixing y and then doing the usual derivative stuff with x!

Partial Derivative

Definition

If there exists a finite limit

$$f_x = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h}$$

then it is called the *partial derivative* of $f(x, y)$ with respect to x, and denoted by f_x or $\frac{\partial f}{\partial x}$.

Partial Derivative

Definition

If there exists a finite limit

$$f_x = \lim_{h \to 0} \frac{f(x+h,y) - f(x,y)}{h}$$

then it is called the *partial derivative* of $f(x,y)$ with respect to x, and denoted by f_x or $\frac{\partial f}{\partial x}$.

Similarly, we define the partial derivative of $f(x,y)$ with respect to y as:

$$f_y = \lim_{h \to 0} \frac{f(x,y+h) - f(x,y)}{h}$$

Partial Derivative

Definition

If there exists a finite limit

$$f_x = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h}$$

then it is called the *partial derivative* of $f(x, y)$ with respect to x, and denoted by f_x or $\frac{\partial f}{\partial x}$.

Similarly, we define the partial derivative of $f(x, y)$ with respect to y as:

$$f_y = \lim_{h \to 0} \frac{f(x, y+h) - f(x, y)}{h}$$

Example

If $f(x, y) = x^2 + y^2$, then:

$$f_x = 2x \qquad \text{and} \qquad f_y = 2y$$

Partial Derivative

The partial derivative is the rate of the function change with respect to only one of the variables, while the others are being kept unchanged (constant).

Partial Derivative

The partial derivative is the rate of the function change with respect to only one of the variables, while the others are being kept unchanged (constant).

So naturally, instead of 1 object to describe the speed, we have 2 objects. We can combine them together in one vector:

Partial Derivative

The partial derivative is the rate of the function change with respect to only one of the variables, while the others are being kept unchanged (constant).

So naturally, instead of 1 object to describe the speed, we have 2 objects. We can combine them together in one vector:

Definition

The vector consisting of the partial derivatives of $f(x, y)$:

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}$$

is called the *gradient* of $f(x, y)$.

In the previous example, $\nabla f = [2x \quad 2y]$.

Partial Derivative

Similarly, for a function of n variables, $f(x_1, \ldots, x_n) = f(\mathbf{x})$ we define partial derivatives as:

$$f_{x_1}(\mathbf{x}) = \frac{\partial f}{\partial x_1}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1 + h, x_2, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{h},$$

$$f_{x_2}(\mathbf{x}) = \frac{\partial f}{\partial x_2}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1, x_2 + h, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{h},$$

$$\vdots$$

$$f_{x_n}(\mathbf{x}) = \frac{\partial f}{\partial x_n}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1, x_2, \ldots, x_n + h) - f(x_1, x_2, \ldots, x_n)}{h}.$$

And the gradient of $f(\mathbf{x})$ as:

$$\nabla f = \frac{df}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} & \frac{\partial f(\mathbf{x})}{\partial x_2} & \ldots & \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{1 \times n}$$

Partial Derivative

As in the single-variable case, partial derivatives have the following

Properties

Partial Derivative

As in the single-variable case, partial derivatives have the following

Properties

1. $$\frac{\partial}{\partial x_i}(f(\mathbf{x}) \pm g(\mathbf{x})) = \frac{\partial f(\mathbf{x})}{\partial x_i} \pm \frac{\partial g(\mathbf{x})}{\partial x_i}$$

Partial Derivative

As in the single-variable case, partial derivatives have the following

Properties

1. $$\frac{\partial}{\partial x_i}(f(\mathbf{x}) \pm g(\mathbf{x})) = \frac{\partial f(\mathbf{x})}{\partial x_i} \pm \frac{\partial g(\mathbf{x})}{\partial x_i}$$

2. $$\frac{\partial}{\partial x_i}(c \cdot f(\mathbf{x})) = c \cdot \frac{\partial f(\mathbf{x})}{\partial x_i}$$

Partial Derivative

As in the single-variable case, partial derivatives have the following

Properties

1. $$\frac{\partial}{\partial x_i}(f(\mathbf{x}) \pm g(\mathbf{x})) = \frac{\partial f(\mathbf{x})}{\partial x_i} \pm \frac{\partial g(\mathbf{x})}{\partial x_i}$$

2. $$\frac{\partial}{\partial x_i}(c \cdot f(\mathbf{x})) = c \cdot \frac{\partial f(\mathbf{x})}{\partial x_i}$$

3. $$\frac{\partial}{\partial x_i}(f(\mathbf{x}) \cdot g(\mathbf{x})) = \frac{\partial f(\mathbf{x})}{\partial x_i} \cdot g(\mathbf{x}) + f(\mathbf{x}) \cdot \frac{\partial g(\mathbf{x})}{\partial x_i}$$

Partial Derivative

Example

Let $f(x, y) = 2x^2$ and $g(x, y) = 4x + 6y$.

$$(f \cdot g)_x = 2x^2(4) + (4x + 6y)(4x) = 24x^2 + 24xy$$
$$(f \cdot g)_y = 2x^2(6) + (4x + 6y)(0) = 12x^2$$

Chain Rule

Seems like the supermarket business is the same old apple stuff?

Not quite right. Sometimes the change of one variable can affect the change of others as well:

Chain Rule

Seems like the supermarket business is the same old apple stuff?

Not quite right. Sometimes the change of one variable can affect the change of others as well:

Example

Assume you're running a supermarket with the profit function

$$f(x, y) = 20x^{1.3} + 30y^{2.1}$$

where x shows how many Duet iced coffees are sold, and y how many Grand Candy iced teas.

Chain Rule

Seems like the supermarket business is the same old apple stuff?

Not quite right. Sometimes the change of one variable can affect the change of others as well:

Example

Assume you're running a supermarket with the profit function

$$f(x, y) = 20x^{1.3} + 30y^{2.1}$$

where x shows how many Duet iced coffees are sold, and y how many Grand Candy iced teas. You notice that obviously, as temperature t gets higher, more people start consuming them. In particular,

$$x(t) = 400 + 3t, \qquad y(t) = 500 + 2t$$

Chain Rule

Seems like the supermarket business is the same old apple stuff?

Not quite right. Sometimes the change of one variable can affect the change of others as well:

Example

Assume you're running a supermarket with the profit function

$$f(x, y) = 20x^{1.3} + 30y^{2.1}$$

where x shows how many Duet iced coffees are sold, and y how many Grand Candy iced teas. You notice that obviously, as temperature t gets higher, more people start consuming them. In particular,

$$x(t) = 400 + 3t, \qquad y(t) = 500 + 2t$$

How does a change of temperature affect your profit?

Chain Rule

In other words,

- if f depends on x and y
- and x (or y) depends on t
- how much does f change as t changes?

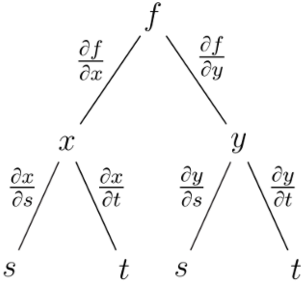
In other words,

- if f depends on x and y
- and x (or y) depends on t
- how much does f change as t changes?

Turns out, there is a simple formula for that:

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial t}$$

which is called the **chain rule**.



Chain Rule

Example

Let $z = \sin(x^2 + y^2),\ x = t^2 + 3,\ y = t^3.$

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = (2x\cos(x^2+y^2)) \cdot (2t) + (2y\cos(x^2+y^2)) \cdot (3t^2)$$
$$= 4xt\cos(x^2+y^2) + 6yt^2\cos(x^2+y^2)$$

Chain Rule

Example

Let $z = \sin(x^2 + y^2),\ x = t^2 + 3,\ y = t^3.$

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = (2x\cos(x^2 + y^2)) \cdot (2t) + (2y\cos(x^2 + y^2)) \cdot (3t^2)$$
$$= 4xt\cos(x^2 + y^2) + 6yt^2\cos(x^2 + y^2)$$

The rule also works for one variable:

Example

Let $z(x) = x^2 + 4x,\ x(t) = 5t^3 + 2t.$

Chain Rule

Example

Let $z = \sin(x^2 + y^2),\ x = t^2 + 3,\ y = t^3.$

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = (2x\cos(x^2 + y^2)) \cdot (2t) + (2y\cos(x^2 + y^2)) \cdot (3t^2)$$
$$= 4xt\cos(x^2 + y^2) + 6yt^2\cos(x^2 + y^2)$$

The rule also works for one variable:

Example

Let $z(x) = x^2 + 4x,\ x(t) = 5t^3 + 2t.$ We can again use the chain rule:

$$\frac{dz}{dt} = \frac{dz}{dx} \cdot \frac{dx}{dt} = (2x + 4) \cdot (15t^2 + 2) = (2 \cdot (5t^3 + 2t) + 4) \cdot (15t^2 + 2)$$

$$= 150t^5 + 80t^3 + 60t^2 + 8t + 8$$

Directional Derivative

When calculating the gradient of a function, we consequently take the rate of change of each coordinate (x_1, x_2, etc), while fixing all other coordinates. What if we change all coordinates simultaneously?

Directional Derivative

When calculating the gradient of a function, we consequently take the rate of change of each coordinate (x_1, x_2, etc), while fixing all other coordinates. What if we change all coordinates simultaneously?

Definition

Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a function, and $\mathbf{v} = [v_1, \ldots, v_n]$ be any vector with $\|\mathbf{v}\| = 1$.

Directional Derivative

When calculating the gradient of a function, we consequently take the rate of change of each coordinate (x_1, x_2, etc), while fixing all other coordinates. What if we change all coordinates simultaneously?

Definition

Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a function, and $\mathbf{v} = [v_1, \ldots, v_n]$ be any vector with $\|\mathbf{v}\| = 1$. If the following limit exists:

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

it is called the *directional derivative* of f *along the vector* $\mathbf{v}$.

Directional Derivative

When calculating the gradient of a function, we consequently take the rate of change of each coordinate (x_1, x_2, etc), while fixing all other coordinates. What if we change all coordinates simultaneously?

Definition

Let $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ be a function, and $\mathbf{v} = [v_1, \ldots, v_n]$ be any vector with $\|\mathbf{v}\| = 1$. If the following limit exists:

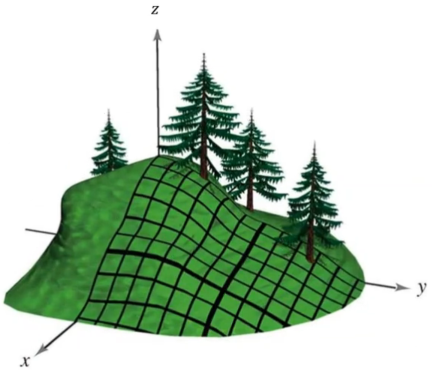
$$\nabla_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}$$

it is called the *directional derivative* of f *along the vector* $\mathbf{v}$.

Note that the directional derivative is a *number* (like partial derivatives), not a vector.

Directional Derivative

The directional derivative shows how much our function changes if we "walk" not only along the x or y-axis, but by an arbitrary direction of our choice.



Directional Derivative

For example, you might want to increase the price of coffee by h drams, but increase the price of tea two times more, i.e. by $2h$ drams.
In this case you would be considering the directional derivative along the vector $[1 \quad 2]$

Directional Derivative

For example, you might want to increase the price of coffee by h drams, but increase the price of tea two times more, i.e. by $2h$ drams.
In this case you would be considering the directional derivative along the vector $[1 \quad 2]$ (divided by $\sqrt{5}$ so its norm is 1).

Directional Derivative

For example, you might want to increase the price of coffee by h drams, but increase the price of tea two times more, i.e. by $2h$ drams.
In this case you would be considering the directional derivative along the vector $[1 \quad 2]$ (divided by $\sqrt{5}$ so its norm is 1).

Question

What happens if you change your variables along the direction $\mathbf{e}_1 = [1 \quad 0]$?
In other words,

$$\nabla_{\mathbf{e}_1} f =$$

Directional Derivative

For example, you might want to increase the price of coffee by h drams, but increase the price of tea two times more, i.e. by $2h$ drams.
In this case you would be considering the directional derivative along the vector $[1 \quad 2]$ (divided by $\sqrt{5}$ so its norm is 1).

Question

What happens if you change your variables along the direction $\mathbf{e}_1 = [1 \quad 0]$?
In other words,

$$\nabla_{\mathbf{e}_1} f = f_x$$

Directional Derivative

How to actually compute the directional derivative?

Directional Derivative

How to actually compute the directional derivative?

Theorem

If $f(\mathbf{x})$ is differentiable at point $\mathbf{x}_0$, then

$$\lim_{h \to 0} \frac{f(\mathbf{x}_0 + h\mathbf{v}) - f(\mathbf{x}_0)}{h} = \nabla f(\mathbf{x}_0) \cdot \mathbf{v}$$

Directional Derivative

How to actually compute the directional derivative?

Theorem

If $f(\mathbf{x})$ is differentiable at point $\mathbf{x}_0$, then

$$\lim_{h \to 0} \frac{f(\mathbf{x}_0 + h\mathbf{v}) - f(\mathbf{x}_0)}{h} = \nabla f(\mathbf{x}_0) \cdot \mathbf{v}$$

- Play with directional derivative

Directional Derivative

How to actually compute the directional derivative?

Theorem

If $f(\mathbf{x})$ is differentiable at point $\mathbf{x}_0$, then

$$\lim_{h \to 0} \frac{f(\mathbf{x}_0 + h\mathbf{v}) - f(\mathbf{x}_0)}{h} = \nabla f(\mathbf{x}_0) \cdot \mathbf{v}$$

- Play with directional derivative

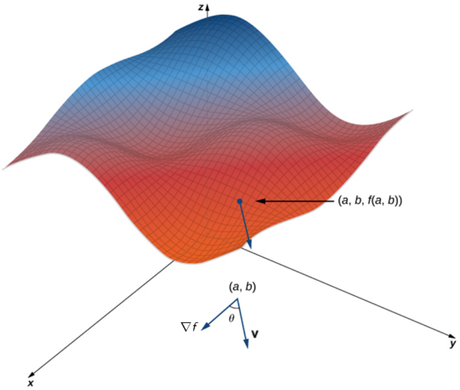
A particularly important question you might ask is:

Question

By which direction should I move, so the function increases the most?

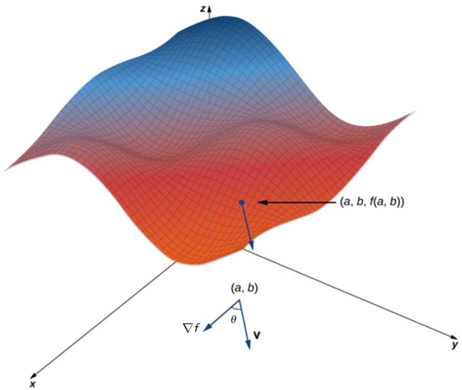
In other words, along which direction does $\nabla_{\mathbf{v}} f$ take its highest value?

Directional Derivative



Suppose $\mathbf{v}$ is any vector (with $\|\mathbf{v}\| = 1$).

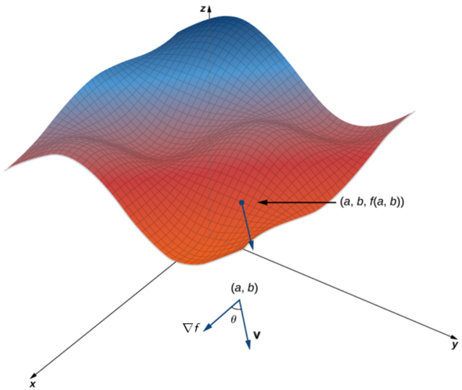
Directional Derivative



Suppose **v** is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} =$$

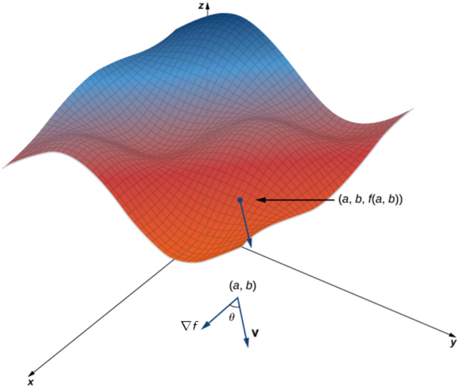
Directional Derivative



Suppose **v** is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos\theta =$$

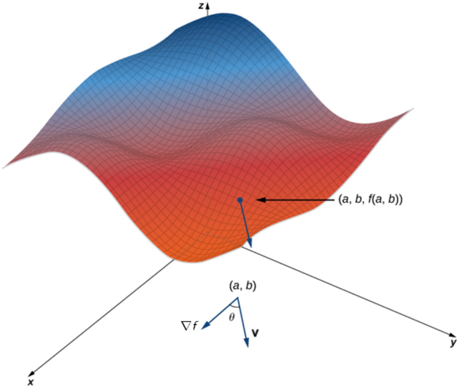
Directional Derivative



Suppose $\mathbf{v}$ is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos\theta = \|\nabla f\| \cos\theta$$

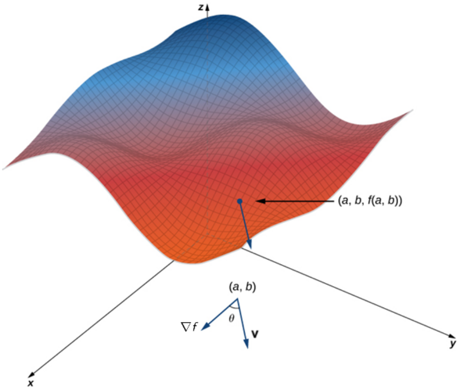
Directional Derivative



Suppose **v** is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos\theta = \|\nabla f\| \cos\theta \leq$$

Directional Derivative

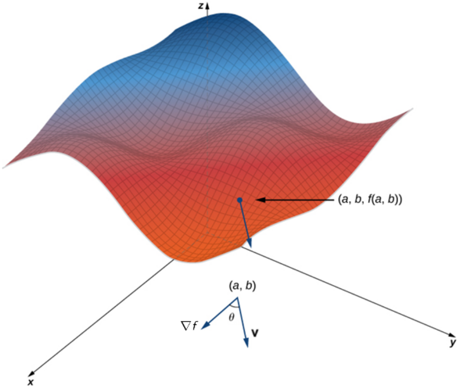


Suppose $\mathbf{v}$ is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos\theta = \|\nabla f\| \cos\theta \leq \|\nabla f\|$$

where θ is the angle between $\mathbf{v}$ and ∇f.

Directional Derivative



Suppose **v** is any vector (with $\|\mathbf{v}\| = 1$). As we saw,

$$\nabla_{\mathbf{v}} f = \nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos\theta = \|\nabla f\| \cos\theta \leq \|\nabla f\|$$

where θ is the angle between **v** and ∇f.

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$
- When does that happen?

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$
- When does that happen?
  When the directions of $\mathbf{v}$ *and* ∇f *coincide*

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$
- When does that happen?
  When the directions of $\mathbf{v}$ *and* ∇f *coincide*
- What does it show?

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$
- When does that happen?
  When the directions of $\mathbf{v}$ *and* ∇f *coincide*
- What does it show?

Theorem

The gradient is the **fastest increasing direction** of the function.

Directional Derivative

$$\nabla_{\mathbf{v}} f = \|\nabla f\| \cos\theta$$

- When does this expression attain its maximum?
  When $\cos\theta = 1$
- When does that happen?
  When the directions of $\mathbf{v}$ *and* ∇f *coincide*
- What does it show?

Theorem

The gradient is the **fastest increasing direction** of the function.

Similarly, $-\nabla f$ is the fastest decreasing direction of the function.

Extrema of a Function

Finally, how can we find the maximum and minimum values of a multivariable function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$?

Extrema of a Function

Finally, how can we find the maximum and minimum values of a multivariable function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$?

Definition

$\mathbf{x}_0$ is called a *local maximum (minimum)* point of f if there exists a positive number $\delta > 0$ such that for all $\mathbf{x}$, if $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, then $f(\mathbf{x}) \leq f(\mathbf{x}_0) \quad (f(\mathbf{x}) \geq f(\mathbf{x}_0))$.

Extrema of a Function

Finally, how can we find the maximum and minimum values of a multivariable function $f(\mathbf{x}) = f(x_1, \dots, x_n)$?

Definition

$\mathbf{x}_0$ is called a *local maximum (minimum)* point of f if there exists a positive number $\delta > 0$ such that for all $\mathbf{x}$, if $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, then $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ $\ (f(\mathbf{x}) \geq f(\mathbf{x}_0))$.

Theorem

If $\mathbf{x}_0$ is a local extremum point of f and there exists $\nabla f(\mathbf{x}_0)$, then $\nabla f(\mathbf{x}_0) = \mathbf{0}$. (The converse is not true).

Extrema of a Function

Finally, how can we find the maximum and minimum values of a multivariable function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$?

Definition

$\mathbf{x}_0$ is called a *local maximum (minimum)* point of f if there exists a positive number $\delta > 0$ such that for all $\mathbf{x}$, if $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, then $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ $(f(\mathbf{x}) \geq f(\mathbf{x}_0))$.

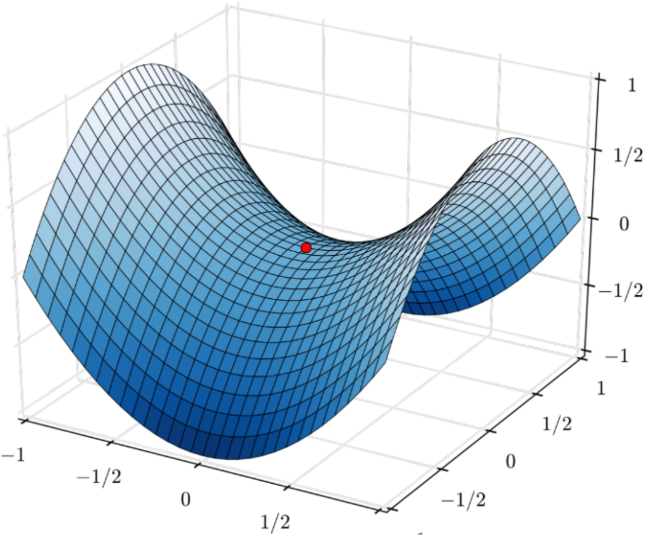
Theorem

If $\mathbf{x}_0$ is a local extremum point of f and there exists $\nabla f(\mathbf{x}_0)$, then $\nabla f(\mathbf{x}_0) = \mathbf{0}$. (The converse is not true).

Definition

$\mathbf{x}_0$ is called a *saddle point* of f if $\nabla f(\mathbf{x}_0) = \mathbf{0}$ but it's not an extremum point.

Extrema of a Function



Extrema of a Function

Question

Which tool do you think we can use to determine whether a critical point is a local extremum or a saddle point?

Extrema of a Function

Question

Which tool do you think we can use to determine whether a critical point is a local extremum or a saddle point?

Second derivatives!

Question

Which tool do you think we can use to determine whether a critical point is a local extremum or a saddle point?

Second derivatives! Except now we have several of them:

- if $f(x, y)$ depends on two variables, we have f_x and f_y

Question

Which tool do you think we can use to determine whether a critical point is a local extremum or a saddle point?

Second derivatives! Except now we have several of them:

- if $f(x, y)$ depends on two variables, we have f_x and f_y
- each of them has two second order derivatives, so in total, we have 4 second order derivatives:

$$\begin{matrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{matrix}$$

Extrema of a Function

Question

Which tool do you think we can use to determine whether a critical point is a local extremum or a saddle point?

Second derivatives! Except now we have several of them:

- if $f(x, y)$ depends on two variables, we have f_x and f_y
- each of them has two second order derivatives, so in total, we have 4 second order derivatives:

$$\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

We combine them in a matrix

Extrema of a Function

Question

Which tool do you think we can use to determine whether a critical point is a local extremum or a saddle point?

Second derivatives! Except now we have several of them:

- if $f(x, y)$ depends on two variables, we have f_x and f_y
- each of them has two second order derivatives, so in total, we have 4 second order derivatives:

$$Hf = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

We combine them in a matrix and call it the *Hessian matrix* of f.

Extrema of a Function

Question

Which tool do you think we can use to determine whether a critical point is a local extremum or a saddle point?

Second derivatives! Except now we have several of them:

- if $f(x, y)$ depends on two variables, we have f_x and f_y
- each of them has two second order derivatives, so in total, we have 4 second order derivatives:

$$Hf = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

We combine them in a matrix and call it the *Hessian matrix* of f.

Just as the gradient plays the role of f' for a multi-variable function, the Hessian matrix plays the role of f''.

Sometimes we even denote the Hessian by $\nabla^2 f$ or $\nabla\nabla f$.

Extrema of a Function

Note that since all second partial derivatives are functions themselves, the Hessian matrix **is a function** as well, i.e. it depends on x and y:

$$Hf(x,y) = \begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix}$$

Extrema of a Function

Note that since all second partial derivatives are functions themselves, the Hessian matrix **is a function** as well, i.e. it depends on x and y:

$$Hf(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}$$

Property

If f_{xy} and f_{yx} are continuous, then they are equal:

$$f_{xy} = f_{yx}$$

In other words, the Hessian matrix is *symmetric.*

Extrema of a Function

Getting back to the question of how to know if a critical point is a local extremum or not, remember how in the single-variable case we had:

Extrema of a Function

Getting back to the question of how to know if a critical point is a local extremum or not, remember how in the single-variable case we had:

Theorem (for one variable)

If $f'(x_0) = 0$ at some point x_0, then:

- if $f''(x_0) > 0$, then x_0 is a local minimum
- if $f''(x_0) < 0$, then x_0 is a local maximum
- if $f''(x_0) = 0$, then we don't know

Extrema of a Function

Getting back to the question of how to know if a critical point is a local extremum or not, remember how in the single-variable case we had:

Theorem (for one variable)

If $f'(x_0) = 0$ at some point x_0, then:

- if $f''(x_0) > 0$, then x_0 is a local minimum
- if $f''(x_0) < 0$, then x_0 is a local maximum
- if $f''(x_0) = 0$, then we don't know

In a very similar manner,

Theorem (for several variables)

If $\nabla f(\mathbf{x}_0) = \mathbf{0}$ at some point $\mathbf{x}_0$, then:

- if $Hf(\mathbf{x}_0) \succ 0$, then $\mathbf{x}_0$ is a local minimum
- if $Hf(\mathbf{x}_0) \prec 0$, then $\mathbf{x}_0$ is a local maximum
- if $Hf(\mathbf{x}_0)$ is not positive/negative definite, then we don't know

Extrema of a Function

In case of two variables, this can be simplified even more.

Extrema of a Function

In case of two variables, this can be simplified even more.

To determine whether a critical point (a, b) is a local extremum or not, we need to calculate two numbers:

- f_{xx}
- and $D = f_{xx}f_{yy} - f_{xy}^2$

at the point (a, b).

Extrema of a Function

In case of two variables, this can be simplified even more.

To determine whether a critical point (a, b) is a local extremum or not, we need to calculate two numbers:

- f_{xx}
- and $D = f_{xx} f_{yy} - f_{xy}^2$

at the point (a, b).

Theorem (for two variables)

If $\nabla f(a, b) = \mathbf{0}$ at some point (a, b), and

- $D > 0$ and $f_{xx} > 0 \quad \Rightarrow \quad$ local minimum
- $D > 0$ and $f_{xx} < 0 \quad \Rightarrow \quad$ local maximum
- $D < 0 \quad \Rightarrow \quad$ saddle point

Extrema of a Function

Example

Let $f(x, y) = x^2 + 3xy - y^2$. To find the critical points, the gradient should be zero.

Extrema of a Function

Example

Let $f(x, y) = x^2 + 3xy - y^2$. To find the critical points, the gradient should be zero.

This is found by solving the system of equations:

$$\begin{aligned} f_x &= 2x + 3y = 0 \\ f_y &= 3x - 2y = 0 \end{aligned}$$

from which we get $x = 0$, $y = 0$.

Extrema of a Function

Example

Let $f(x, y) = x^2 + 3xy - y^2$. To find the critical points, the gradient should be zero.

This is found by solving the system of equations:

$$f_x = 2x + 3y = 0$$
$$f_y = 3x - 2y = 0$$

from which we get $x = 0$, $y = 0$.

Now, compute the second partial derivatives at the point $(0, 0)$:

$$f_{xx} = 2, \quad f_{yy} = -2, \quad f_{xy} = 3$$

Extrema of a Function

Example

Let $f(x, y) = x^2 + 3xy - y^2$. To find the critical points, the gradient should be zero.

This is found by solving the system of equations:

$$f_x = 2x + 3y = 0$$
$$f_y = 3x - 2y = 0$$

from which we get $x = 0$, $y = 0$.

Now, compute the second partial derivatives at the point $(0, 0)$:

$$f_{xx} = 2, \quad f_{yy} = -2, \quad f_{xy} = 3$$

Then, $D = f_{xx} f_{yy} - f_{xy}^2 = 2 \cdot (-2) - 3^2 = -4 - 9 = -13 < 0$.

Extrema of a Function

Example

Let $f(x, y) = x^2 + 3xy - y^2$. To find the critical points, the gradient should be zero.

This is found by solving the system of equations:

$$f_x = 2x + 3y = 0$$
$$f_y = 3x - 2y = 0$$

from which we get $x = 0$, $y = 0$.

Now, compute the second partial derivatives at the point $(0, 0)$:

$$f_{xx} = 2, \quad f_{yy} = -2, \quad f_{xy} = 3$$

Then, $D = f_{xx}f_{yy} - f_{xy}^2 = 2 \cdot (-2) - 3^2 = -4 - 9 = -13 < 0$.

Since $D < 0$, $(0, 0)$ is a saddle point.

Taylor Expansion (optional)

Just as in the single-variable case, we can approximate a multi-variable function $f(\mathbf{x})$ around some point $\mathbf{a}$ using its derivatives at that point.

Taylor Expansion (optional)

Just as in the single-variable case, we can approximate a multi-variable function $f(\mathbf{x})$ around some point $\mathbf{a}$ using its derivatives at that point.

In case of one variable, we had the linear approximation:

$$f(x) \approx f(a) + f'(a)(x - a)$$

Taylor Expansion (optional)

Just as in the single-variable case, we can approximate a multi-variable function $f(\mathbf{x})$ around some point $\mathbf{a}$ using its derivatives at that point.

In case of one variable, we had the linear approximation:

$$f(x) \approx f(a) + f'(a)(x - a)$$

which was a *line* tangent to the graph of f at the point a.

In case of several variables, this becomes a *plane* tangent to the surface of f at the point $\mathbf{a}$:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

Taylor Expansion (optional)

As for the quadratic approximation, in one variable we had:

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2$$

Taylor Expansion (optional)

As for the quadratic approximation, in one variable we had:

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2$$

or, equivalently,

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{1}{2}(x-a)f''(a)(x-a)$$

Taylor Expansion (optional)

As for the quadratic approximation, in one variable we had:

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2$$

or, equivalently,

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{1}{2}(x-a)f''(a)(x-a)$$

In several variables, this becomes a *quadric surface:*

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})$$

Taylor Expansion (optional)

As for the quadratic approximation, in one variable we had:

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2$$

or, equivalently,

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{1}{2}(x-a)f''(a)(x-a)$$

In several variables, this becomes a *quadric surface:*

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})$$

This surface is not only tangent to the surface of f at $\mathbf{a}$, but it also has the same convexity/concavity as that surface.

Taylor Expansion (optional)

As for the quadratic approximation, in one variable we had:

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2$$

or, equivalently,

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{1}{2}(x-a)f''(a)(x-a)$$

In several variables, this becomes a *quadric surface:*

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a}) + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})$$

This surface is not only tangent to the surface of f at $\mathbf{a}$, but it also has the same convexity/concavity as that surface.

The Taylor expansion can be extended to even higher orders, but we won't need that – instead check some examples.

Jacobian Matrix (optional)

Finally, if we have a vector-valued function, i.e.

$$\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$$

which takes a vector and returns a vector:

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$$

Jacobian Matrix (optional)

Finally, if we have a vector-valued function, i.e.

$$\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$$

which takes a vector and returns a vector:

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$$

then if we compute the gradients of each component function, we will get m vectors of size n each.

Jacobian Matrix (optional)

Finally, if we have a vector-valued function, i.e.

$$\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$$

which takes a vector and returns a vector:

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{bmatrix}$$

then if we compute the gradients of each component function, we will get m vectors of size n each.

The resulting matrix is called the *Jacobian matrix* of $\mathbf{f}$:

$$J\mathbf{f}(\mathbf{x}) = \begin{bmatrix} \nabla f_1(\mathbf{x}) \\ \nabla f_2(\mathbf{x}) \\ \vdots \\ \nabla f_m(\mathbf{x}) \end{bmatrix} \in \mathbb{R}^{m \times n}$$