1 Fri 9/18 11.3: Partial Derivatives

Main idea: fix all except one variables and differentiate with respect to that one variable.

Example 1.1

Say $f(x,y,z) = x^2yz + z^2$, a function from $\mathbb{R}^3$ to $\mathbb{R}$. Then

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(x^2yz + z^2) = f_x(x,y,z) = D_x f(x,y,z) = 2xyz$$

Lots of notations, but ∂ is definitely the most elegant one!

Likewise we have

$$\frac{\partial f}{\partial y} = x^2 z \text{ and } \frac{\partial f}{\partial z} = x^2 y + 2z.$$

In general, if $f : \mathbb{R}^n \to \mathbb{R}^m$, $\dfrac{\partial \mathbf{F}}{\partial x_i}$ is a vector valued functions of $(x_1, \ldots, x_n)$, and the principle of holding all but one variables still holds — partial differentiate each component.

2 Mon 9/21

Geometric meaning of $f(x,y) : \mathbb{R}^2 \to \mathbb{R}$? We can draw the *graph* of these functions as surfaces in 3d space. It still passes the vertical line test — vertical along the z-axis.

Example 2.1

Consider the graph of $f(x,y) = x^2 + 2y^2$. The graph is an elliptic paraboloid. What about partial derivatives?

$$\frac{\partial f}{\partial x}(x,y) = 2x \text{ and } \frac{\partial f}{\partial y}(x,y) = 4y$$

Consider $(x,y) = (1,1)$. This gives $(1,1,f(1,1)) = (1,1,3)$ on the surface. Then

$$\frac{\partial f}{\partial x}(1,1) = 2 \text{ and } \frac{\partial f}{\partial y}(1,1) = 4$$

Think of $\dfrac{\partial f}{\partial x}(1,1)$ as intersecting the graph of f (the surface) with the plane parallel to the xz-plane with y held to be a constant with value 1. "How fast is $f(x,y)$ changing with respect to x when we fix y at this specific (x_0, y_0)?"

Likewise, the y-partial address the problem "how fast is $f(x,y)$ changing with respect to y when we fix x at this specific (x_0, y_0)?"

Definition 1

Higher partial derivatives: some examples are $\frac{\partial^2 f}{\partial x^2} f$ and $f_{xy} = \frac{\partial}{\partial y}\frac{\partial}{\partial x} f = \frac{\partial}{\partial x}\frac{\partial}{\partial y} f = \frac{\partial^2}{\partial x \partial y} f$. Ordering does not matter.

Example 2.2

Consider $f(x,y) = e^{2xy} + y^3$. Then

$$\frac{\partial f}{\partial x}(x,y) = 2ye^{2xy} \text{ and } \frac{\partial f}{\partial y}(x,y) = 2xe^{2xy} + 3y^2 \text{ are called first-order partial derivatives.}$$

Now consider higher-order partial derivatives:

$$\frac{\partial^2 f}{\partial x^2}(x,y) = 4y^2 e^{2xy} \text{ and } \frac{\partial^2 f}{\partial y^2}(x,y) = 4x^2 e^{2xy} + 6y$$
$$\frac{\partial^2 f}{\partial y \partial x}(x,y) = 2e^{2xy} + 4xye^{2xy} \text{ and } \frac{\partial^2 f}{\partial x \partial y} = 2e^{2xy} + 4xye^{2xy}$$

The two on the first line are called **second-order partial derivatives** and the two on the last line are called **mixed partial derivatives**.

Remark: Way beyond scope!

If the second order mixed partials of f are continuous functions, then they are equal. Likewise if the n^{th}-order partials are continuous functions, then the ordering doesn't matter for n^{th}-order mixed partials.

Implicit Differentiation

Example 2.3

Consider the surface in $\mathbb{R}^3$ defined by the equation

$$\sin x + \sin(yz) - z = 0 \text{ (a level set description)}$$

This may be something unlike what we've considered as it cannot be written as $z = f(x,y)$, but the idea here is that we can view this as a graph of $f(x,y)$ near "most" points.

Upshot: if we have $f(x,y)$ defined *implicitly* like this, starting from an equation (level set of surface), we can still use the equation to help compute the partial derivatives.

3 Wed 9/23 Wrap up of 11.3 and intro to 11.4

Implicit differentiation from last time: an equation like $\sin x + \sin(yz) - z = 0$ defines (near most (x, y, z)) the value of z implicitly as a function of x, y. (Could use $f(x, z)$ or $f(y, z)$ to define y or x, respectively, as well.)

Problem 1

How to compute $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$?

Solution

We know that

$$\sin(x) + \sin(yf(x, y)) - f(x, y) = 0,$$

and if we take $\dfrac{\partial}{\partial x}$ on both sides, we still get a true equation:

$$\cos(x) + \cos(yf(x, y))y\frac{\partial f}{\partial x}(x, y) - \frac{\partial f}{\partial x}(x, y) = 0.$$

Now f is still inside the cosine function but partial derivatives are outside. Then

$$\frac{\partial f}{\partial x}(x, y) \cdot (y\cos(yf(x, y)) - 1) = -\cos x \implies \frac{\partial f}{\partial x}(x, y) = \frac{\cos x}{1 - y\cos(yf(x, y))}$$

For a shorthand notation, write z instead of $f(x, y)$ or f everywhere. However, unlike when usually computing partial derivatives, z should be treated as a function of x and y rather than a constant. Then the cumbersome $\dfrac{\partial f}{\partial x}(x, y)$ becomes $\dfrac{\partial z}{\partial x}$. In this example:

$$\begin{aligned}
\sin x + \sin(yz) - z &= 0 \\
\cos x + \cos(yz)y\frac{\partial z}{\partial x} - 1 &= 0 \\
\frac{\partial z}{\partial x} &= \frac{\cos(x)}{1 - y\cos(yz)}
\end{aligned}$$

and for $\dfrac{\partial z}{\partial y}$,

$$\begin{aligned}
\sin x + \sin(yz) - z &= 0 \\
0 + \cos(yz)\left[z + y\frac{\partial z}{\partial y}\right] - \frac{\partial z}{\partial y} &= 0 \\
\frac{\partial z}{\partial y}(y\cos(yz) - 1) &= -z\cos(yz) \\
\frac{\partial z}{\partial y} &= \frac{z\cos(yz)}{1 - y\cos yz}
\end{aligned}$$

A bit about Partial Differential Equations, PDEs

These equations describe things over multiple directions (spatial), or $\geqslant 1$ spatial direction plus time.

Example 3.1

Two spacial directions: ind functions $u(x, y)$, $u : \mathbb{R}^2 \to \mathbb{R}$, satisfying

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad \text{(Laplace Equation)}$$

Or consider one spacial equation with one time direction: find functions $u(x, t)$ satisfying

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} \quad \text{(wave equation)}$$

or

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} \quad \text{(heat equation)}$$

11.4 Tangent Planes, Linear Approximations

Problem 2

Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = x^2 + y^2$. What what's the equation of the tangent plane to the surface at the point $(x_0, y_0, f(x_0, y_0))$?

Solution

Call $z_0 = f(x_0, y_0)$. We have a point (x_0, y_0, z_0) on the plane. All that remains is to find a normal vector. Claim:

$$\left\langle \frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0), -1 \right\rangle$$

is a normal vector to the tangent plane. Therefore the scalar equation of the plane is

$$\frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) = z - z_0.$$

If $(x_0, y_0, z_0) = (1, 1, 2)$, the tangent plane becomes

$$z - 2 = 2(x - 1) + 2(y - 1) \implies 2x + 2y - z = 2.$$

Next time we will explain why the vector above is a normal vector to the tangent plane using the concept of gradient from 11.6.

4 Mon 9/28 11.6 Gradient

From last time: why is

$$\left\langle \frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0), -1 \right\rangle$$

the normal vector to the tangent plane at $f(x_0, y_0)$?

Note: the book uses 11.3 for partial derivatives, but (Manion thinks) it's clearer with gradient.

First: the surface S is the graph of f but it's also the level set 0 of the function

$$F(x, y, z) = f(x, y) - z.$$

The most natural way to get normal vectors to level sets of functions $F : \mathbb{R}^n \to \mathbb{R}$ is to take gradient vector of F at the point.

Definition 2

Let $F : \mathbb{R}^n \to \mathbb{R}$ be a continuous function. It has n first-order partial derivatives at a point p in $\mathbb{R}^n$:

$$\frac{\partial F}{\partial x_1}(p), \frac{\partial F}{\partial x_2}(p), \dots, \frac{\partial F}{\partial x_n}(p).$$

We define the **gradient vector** of $F : \mathbb{R}^n \to \mathbb{R}$ (a **scalar-valued function**) at point $p = (x_1 \dots, x_n)$ as

$$(\nabla F)(x_1, x_2, \dots, x_n) = \left\langle \frac{\partial F}{\partial x_1}(x_1, \dots, x_n), \frac{\partial F}{\partial x_2}(x_1, \dots, x_n), \dots, \frac{\partial F}{\partial x_n}(x_1, \dots, x_n) \right\rangle.$$

Remark

If $f : \mathbb{R}^n \to \mathbb{R}^m$ we have *Jacobian matrix* instead of gradient, and we'll get to Jacobian very soon.

Note that ∇F takes input from $\mathbb{R}^n$ and returns an output in $\mathbb{R}^n$ again. $\nabla F : \mathbb{R}^n \to \mathbb{R}^n$.

A new way to visualize a function: a vector field in $\mathbb{R}^n$ and a sequence (p_n) of points and the drawing the vectors rooted at each point.

Example 4.1

Let $F : \mathbb{R}^2 \to \mathbb{R}$ defined by $F(x, y) = x^2 + 2y^2$.

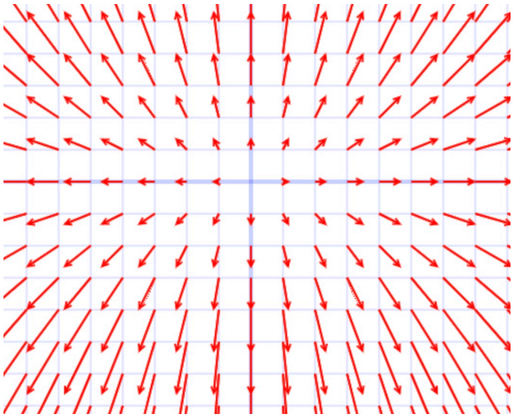
(1) What's ∇F?

(2) Draw ∇F as a vector field on 2d space.

Solution

(1) $\nabla F = \langle 2x, 4y \rangle$.

(2) See below.



Remark

If we draw the level sets of F, the gradient will be normal to the level sets everywhere. More importantly,

(1) $(\nabla F)(p)$ points in the direction of *maximum* increase of F and

(2) $(\nabla F)(p)$ are perpendicular to "directions where F is constant", tangent directions.

We'll see both soon.

Going back to the example of $f(x,y) = 0$ or the level set of $F(x,y,z) := f(x,y) - z$ at level 0:
How to get a normal vector to tangent plane of this level set? Answer:

$$\begin{aligned}(\nabla F)(x_0,y_0,z_0) &= \left\langle \frac{\partial F}{\partial x}(x_0,y_0,z_0), \frac{\partial F}{\partial y}(x_0,y_0,z_0), \frac{\partial F}{\partial z}(x_0,y_0,z_0) \right\rangle \\ &= \left\langle \frac{\partial f}{\partial x}(x_0,y_0), \frac{\partial f}{\partial y}(x_0,y_0), -1 \right\rangle\end{aligned}$$

Remark

We can use ∇F (or Jacobian more generally) not just to understand tangent planes, but also to *approximate F linearly.*

Example 4.2

Consider $f : \mathbb{R}^2 \to \mathbb{R}$ ($f(x, y)$ is a scalar). Then

(1) The tangent plane to the graph of f at (x_0, y_0, z_0) is also the graph of some function!

(2) It is the graph of

$$L(x, y) = z_0 + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(y - y_0).$$

(3) Since graphs of f and L are close near (x_0, y_0), we may use $L(x, y)$, a linear function, to approximate $f(x, y)$ near (x_0, y_0).

Definition 3

The real definition of **total differentiability** (not just partial) is that F is differentiable if and only if

$$\lim_{(x,y)\to(0,0)} \frac{F(x, y) - L(x, y)}{\| \langle x, y \rangle \|} = 0.$$

Analogous to the single-variable ordinary differentiability if and only if:

$$\lim_{t\to t_0} f(t) - f'(t)(t - t_0) = 0.$$

Problem 3

Linearly approximate $f(0.1, 1.1)$ where $f(x, y) = y/(1 + e^x)$.

Solution

This point is close to $f(0, 1)$. All we need to do is to compute

$$f(0.1, 1.1) \approx f(0, 1) + (0.1 - 0)\frac{\partial f}{\partial x}(0, 1) + (1.1 - 1)\frac{\partial f}{\partial y}(0, 1) = f(0, 1) + 0.1\frac{\partial f}{\partial x}(0, 1) + 0.1\frac{\partial f}{\partial y}(0, 1).$$

Actual computation omitted.

Remark

Think of this as “how much would f increase as x increases by $(x - x_0)$ and y increases by $(y - y_0)$? The answer would be

$$(x - x_0)\frac{\partial f}{\partial x}(x_0, y_0) + (y - y_0)\frac{\partial f}{\partial y}(x_0, y_0)$$

where the first part describes “how much f increases as x increases by $(x - x_0)$” and the second describes “how much f increases as y increases by $(y - y_0)$". Of course this function only has two variables, but that should give a fairly accurate approximation for (x, y) near (x_0, y_0).

5 Wed 9/30 11.4/5/6 Jacobian, Chain Rule, Direc. Deriv.

From last time: note that we can write

$$L(x,y) = z_0 + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

as the dot product

$$L(x,y) = z_0 + (\nabla F)(x_0, y_0) \cdot \langle x - x_0, y - y_0 \rangle .$$

How about $F : \mathbb{R}^n \to \mathbb{R}^m$? When $m = 1$ we get a scalar involving a dot product, whereas for $m > 1$ we get a *vector* where each entry involves a scalar represented by dot product.

For the degenerate case $m = 1$, we can write the linear approximation as

$$L(x,y) = [z_0] + \begin{bmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}.$$

For the general case $F : \mathbb{R}^n \to \mathbb{R}^m$, we get a matrix $J_F(x_1, \dots, x_n)$ at each point $(x_1, \dots, x_n)$ of $\mathbb{R}^n$ rather than just a vector ∇F. For example, $F : \mathbb{R}^3 \to \mathbb{R}^2$ think of it as $[F_1(x,y,z), F_2(x,y,z)]$.

Definition 4: Jacobian

If $F : \mathbb{R}^3 \to \mathbb{R}^2$, the **Jacobian** at (x, y, z) is

$$J_F(x,y,z) = \begin{bmatrix} \frac{\partial F_1}{\partial x}(x,y,z) & \frac{\partial F_1}{\partial y}(x,y,z) & \frac{\partial F_1}{\partial z}(x,y,z) \\ \frac{\partial F_1}{\partial x}(x,y,z) & \frac{\partial F_2}{\partial y}(x,y,z) & \frac{\partial F_2}{\partial z}(x,y,z) \end{bmatrix}.$$

Of course it works with any m, n!

Example 5.1

The Jacobian of $F(x, y, z)$ defined by $(x, y, z) \mapsto (x^2 y + xyz, 3z - 2xy^2)$ is

$$J_F(x,y,z) = \begin{bmatrix} 2xy + yz & x^2 + xz & xy \\ -2y^2 & -4xy & 3 \end{bmatrix}.$$

Example 5.2

Two degenerate cases:

If $F(t): t \mapsto (t, t^2, t^3)$, then

$$J_F(t) = \begin{bmatrix} 1 \\ 2t \\ 3t^2 \end{bmatrix}, \text{ a column vector!}$$

If $F(x, y, z) = x^2yz$, then

$$J_F(x, y, z) = \begin{bmatrix} 2xyz & x^2z & x^2y \end{bmatrix}, \text{ a row vector, the same as } \nabla F.$$

Definition 5: Linear Approximation using Jacobian

If $F : \mathbb{R}^n \to \mathbb{R}^m$ and $F(\vec{x_0}) = \vec{y_0}$ (note that $\vec{x_0} \in \mathbb{R}^n, \vec{y_0} \in \mathbb{R}^m$), then

$$\vec{L}(\vec{x}) = \vec{y_0} + (J_F(\vec{x_0})(\vec{x} - \vec{x_0}).$$

Example 5.3

Let $F : \mathbb{R}^3 \to \mathbb{R}^2$. Define $F : (x, y, z) \mapsto (u, v)$. Let $F(x_0, y_0, z_0) = (u_0, v_0)$, then the linear approximation $\vec{L}(x, y, z)$ near (x_0, y_0, z_0) is given by

$$\vec{L}(x,y,z) = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} \frac{\partial F_1}{\partial x}(x_0,y_0,z_0) & \frac{\partial F_1}{\partial y}(x_0,y_0,z_0) & \frac{\partial F_1}{\partial z}(x_0,y_0,z_0) \\ \frac{\partial F_2}{\partial x}(x_0,y_0,z_0) & \frac{\partial F_2}{\partial y}(x_0,y_0,z_0) & \frac{\partial F_2}{\partial z}(x_0,y_0,z_0) \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix} = \begin{bmatrix} \text{First in } \mathbb{R}^2 \\ \text{Second in } \mathbb{R}^2 \end{bmatrix}.$$

Notice how this is the direct generalization of

$$L(x,y) = z_0 + \begin{bmatrix} \frac{\partial f}{\partial x}(x_0,y_0) & \frac{\partial f}{\partial x}(x_0,y_0) \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}$$

from our previous $\mathbb{R}^2 \to \mathbb{R}$ example.

Definition 6

A function $\mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\vec{x_0} \in \mathbb{R}^n$) with derivative J (Jacobian) if

$$\lim_{\vec{x} \to \vec{x_0}} F(\vec{x}) - \frac{F(\vec{x_0}) + J(\vec{x} - \vec{x_0})}{\|\vec{x} - \vec{x_0}\|} = 0.$$

Theorem 7: Chain Rule

For $F:\mathbb{R}^n \to \mathbb{R}^m$ and $G:\mathbb{R}^m \to \mathbb{R}^k$,

$$J_{G\circ F}(\vec{x}) = J_G(F(\vec{x}))J_F(\vec{x})$$

if the functions are "nice" enough (the definition above holds).

6 Fri 10/2

Continuing from last time: "the linear approximation to a composite function is the composite of linear approximations by the chain rule".

Consider a special case of chain rule:

$$f:\mathbb{R}^2 \to \mathbb{R} \text{ and } G:\mathbb{R} \to \mathbb{R}^2$$

with $G(t) = (g, h)$. We have

$$J_G(t) = \begin{bmatrix} \frac{\mathrm{d}g}{\mathrm{d}t}(t) \\ \frac{\mathrm{d}h}{\mathrm{d}t}(t) \end{bmatrix} \text{ and } J_f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial x}(x,y) & \frac{\partial f}{\partial y}(x,y) \end{bmatrix}.$$

Then, by chain rule,

$$J_{f\circ G}(t) = \begin{bmatrix} \frac{\partial f}{\partial x}(g(t),h(t)) & \frac{\partial f}{\partial y}(g(t),h(t)) \end{bmatrix} \begin{bmatrix} \frac{\mathrm{d}g}{\mathrm{d}t}(t) \\ \frac{\mathrm{d}h}{\mathrm{d}t}(t) \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x}(g(t),h(t))\frac{\mathrm{d}g}{\mathrm{d}t}(t) + \frac{\partial f}{\partial y}(g(t),h(t))\frac{\mathrm{d}h}{\mathrm{d}t}(t) \end{bmatrix},$$

a constant.

In the book's notation, if we denote $g(t)$ as x and $h(t)$ as y, then

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\partial z}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial z}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}.$$

Example 6.1

Let $f(x,y) = x^2e^{xy}, x(t) = \cos t, y(t) = \sin t$, then if we define $G(t) = (\cos t, \sin t)$, by the Jacobian,

$$J_f(x,y) = \begin{bmatrix} 2xe^{xy} + x^2ye^{xy} & x^3e^{xy} \end{bmatrix} \text{ and } J_G(t) = \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix}$$

and

$$\begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}t}(f(x(t),y(t))) \end{bmatrix} = J_{f\circ G}(t) = \begin{bmatrix} 2\cos(t)e^{\cos t\sin t} + \cos^2 t\sin te^{\cos t\sin t} & \cos^3 te^{\cos t\sin t} \end{bmatrix} \begin{bmatrix} -\sin t \\ \cos t \end{bmatrix}$$

and whatever that equals to is the solution.

Example 6.2

Another special case of chain rule: define $G : \mathbb{R}^2 \to \mathbb{R}^2$ by $(s,t) \mapsto (g(s,t), h(s,t)$ and $f : \mathbb{R}^2 \to \mathbb{R}$. Then

$$J_G(s,t) = \begin{bmatrix} \frac{\partial g}{\partial s}(s,t) & \frac{\partial g}{\partial t}(s,t) \\ \frac{\partial h}{\partial s}(s,t) & \frac{\partial h}{\partial t}(s,t) \end{bmatrix} \text{ and } J_f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial x}(x,y) & \frac{\partial f}{\partial y}(x,y) \end{bmatrix}.$$

Then

$$J_{f\circ G}(s,t) = J_f(G(s,t))J_G(s,t) = \begin{bmatrix} \frac{\partial f}{\partial x}(g(s,t),h(s,t)) & \frac{\partial f}{\partial y}(g(s,t),h(s,t)) \end{bmatrix} \begin{bmatrix} \frac{\partial g}{\partial s}(s,t) & \frac{\partial g}{\partial t}(s,t) \\ \frac{\partial h}{\partial s}(s,t) & \frac{\partial h}{\partial t}(s,t) \end{bmatrix}$$

In the book, if we write $f(x,y)$ as $z(x,y)$, $g(s,t)$ as $x(s,t)$ and $h(s,t)$ as $y(s,t)$, then

$$\frac{\partial s}{\partial x} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \text{ and } \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

7 Mon 10/5

Example 7.1

Solving an implicit differentiation

$$F(x,y,f(x,y)) = 0 :$$

Differentiating both sides gives

$$\begin{bmatrix} \frac{\partial F}{\partial x}(x,y,z) & \frac{\partial F}{\partial y}(x,,z) & \frac{\partial F}{\partial z}(x,y,z) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{\partial f}{\partial x}(x,y) & \frac{\partial f}{\partial y}(x,y) \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

where the first matrix is $J_F(x,y,z)$ and the second $J_{(x,y,f(x,y))}(x,y)$. The RHS is, of course, derivative of 0 with respect to both variables. i

This gives

$$\frac{\partial f}{\partial x}(x,y) = -\frac{\frac{\partial F}{\partial x}(x,y,z)}{\frac{\partial F}{\partial z}(x,y,z)} \text{ and } \frac{\partial f}{\partial x}(x,y) = -\frac{\frac{\partial F}{\partial y}(x,y,z)}{\frac{\partial F}{\partial z}(x,y,z)}$$

11.6 Directional Derivatives

Given $F : \mathbb{R}^n \to \mathbb{R}^m$, the partial derivative $\frac{\partial F}{\partial x_i}$ looks at how F changes along lines parallel to the x_i-axis. For more general lines, we introduce the following:

Definition 8: Directional derivative

Given $f : \mathbb{R}^n \to \mathbb{R}^m$, the **directional derivative** of F at $p \in \mathbb{R}^n$ in the direction of $\mathbf{u}$, any unit vector, denoted as $D_{\mathbf{u}}F(p)$, is defined as

$$D_{\mathbf{u}}F(p) := \lim_{t\to 0} \frac{F(p + t\mathbf{u}) - F(p)}{t}.$$

Consider $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G(t) : \mathbb{R} \to \mathbb{R}^n$ defined by $G(t) = p + t\mathbf{u}$. Then

$$(F \circ G)'(t) = \lim_{t\to 0} \frac{F(G(t)) - F(G(0))}{t} = \lim_{t\to 0} \frac{F(p + t\mathbf{u}) - F(p)}{t}.$$

Chain rule gives

$$J_F(p)J_G(0) = J_F(p)\mathbf{u}.$$

If $m = 1$, then $J_F(p)$, the gradient vector, is merely a row vector. Hence if $F : \mathbb{R}^n \to \mathbb{R}$, the directional derivative becomes $\nabla F(p) \cdot \mathbf{u}$, a dot product.

Since

$$\nabla F(p) \cdot \mathbf{u} = |\nabla F(p)||\mathbf{u}| \cos\theta,$$

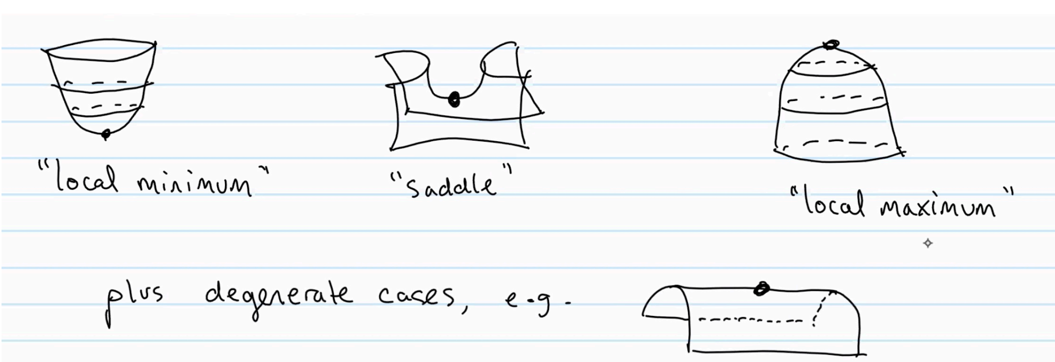
we see that the directional derivative achieves maximum along the direction of the gradient, and the maximum is simply $|\nabla F(p)|$. On the other hand, $\nabla F(p)$ is perpendicular to the level set of F at level $F(p)$, and this plane consists of all vectors orthogonal to $\nabla F(p)$.

8 Wed 10/7

Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$. Then the graph is a surface in $\mathbb{R}^3$. The analogue of $f'(x) = 0$ for a single-variable function is $\nabla F(x, y) = \langle 0, 0\rangle$. For a more general function, $f : \mathbb{R}^n \to \mathbb{R}^m$, the "critical points" are the points at which the Jacobian vanishes.

In single-variable calculus, these critical points correspond to either a degenerate case where $f''(x) = 0$ or a *local minimum / maximum*.

For $f : \mathbb{R}^2 \to \mathbb{R}$, there are three nondegenerate things that can happen to the graph of f: local minimum, a saddle point, or a local maximum.



In general, for $f : \mathbb{R}^n \to \mathbb{R}$, there are $n+1$ different types of critical points: *"morse index* 0 *critical points, ..., morse index* n *critical points".*

How to tell apart different types of critical points? This involves second derivatives + linear algebra.

The Jacobian matrix encodes "total" first derivative of a multivariable function $f : \mathbb{R}^n \to \mathbb{R}^m$. We will focus on the case $\mathbb{R}^n \to \mathbb{R}$ here, where the Jacobian is merely ∇F. Then the first derivative is

$$J_f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right]$$

where we can view ∇f as a function $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$. Then Jacobian gives the "second derivative" of f :

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial^2 x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

This is called the **Hessian Matrix**.

Example 8.1

If $f : \mathbb{R} \to \mathbb{R}$, H_f is simply $\left[\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}\right]$, the second derivative.

If $f : \mathbb{R}^2 \to \mathbb{R}$, then

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

Example 8.2

Let $f(x,y) = xy - 2x - 2y - x^2 - y^2$.

(1) Find critical points of f.

(2) Compute $H_f(x,y)$.

Solution

(1) First note that $\nabla(x,y) = \langle y-2-2x, x-2-2y\rangle$. The critical points correspond to when $\nabla f = \langle 0,0\rangle$. This gives

$$\begin{cases} y-2x=2 \\ x-2y=2 \end{cases} \implies \begin{cases} x=-2 \\ y=-2 \end{cases}$$

From this we see that the only critical point is $(x,y) = (-2,-2)$.

(2)

$$H_f(x,y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}.$$

Remark

For the "second derivative test", do linear algebra on H_f and take eigenvalues. Basic facts / definitions:

(1) "non-degenerate" critical points are defined to be points where H_f has nonzero determinant, i.e., invertible.

(2) The more index of non-degenerate critical points is the number of negative eigenvalues of H_f at the critical point.

(3) Local min: morse index 0, no negative λ's.

(4) Local max: morse index n, all negative λ's.

9 Fri 10/9

Suppose $f : \mathbb{R}^2 \to \mathbb{R}$, then

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

(1) If $\det(H_f) = 0$ then the second-derivative test is inconclusive.

(2) If $\det(H_f) < 0$ then H_f has one positive eigenvalue and one negative eigenvalue, corresponding to a saddle point.

(3) If $\det(H_f) > 0$ then H_f has eigenvalues of the same sign. If this is the case, we either have a local min or a local max.

(1) If f_{xx} (or f_{yy}) is positive then (x,y) is a local min.

(2) Otherwise, if both are negative then (x,y) is a local max.

Example 9.1

Consider $f(x,y) = xy - 2x - 2y - x^2 - y^2$ from last time. We know f has a unique critical point at $(-2,-2)$ and

$$H_f \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \text{ and } \det(H_f) = 3.$$

Nonzero determinant along with negative diagonal entries imply that (x,y) is a local max.

Global Minima and Maxima

Theorem 9: Extreme Value Theorem

They sometimes exist and sometimes don't exist. However, if we consider $f : D \to \mathbb{R}$ rather than $f : \mathbb{R}^2 \to \mathbb{R}$, where D is a closed and bounded subset of $\mathbb{R}^2$, then f attains global maximum and minimum.

Remark

This is exactly what 425a covered. By Heine-Borel, D is closed and bounded in $\mathbb{R}^2$ so it is compact. Then the continuous image of a compact set is compact and therefore is closed and bounded. Hence $f(D)$ attains all its limits, including the maximum and minimum.

Solution

Approach:

(1) Think of D in terms of boundary and interior.

(2) For interior, compute ∇F, set it equal to 0, and solve for x, y. *Compute* the value of f at these points.

(3) Then analyze the values along the boundary. Keep in mind that the points on boundary could well be absolute maxima/minima without being critical points.

Example 9.2

Let $D = \{(x,y) \in \mathbb{R}^2 \mid |x| \leqslant 1, |y| \leqslant 1\}$. Let $f : D \to \mathbb{R}$ be $(x,y) \mapsto x^2 + y^2 + x^2y + 4$. Find the maximum and minimum values of f on D.

Solution

The interior of D is $(-1,1) \times (-1,1)$. Since $\nabla F(x,y) = \left\langle 2x + 2xy, 2y + x^2 \right\rangle$. Setting it to $\langle 0,0 \rangle$ gives

$$\begin{cases} 2x + 2xy = 0 \\ 2y + x^2 = 0 \end{cases} \implies \begin{cases} 2x(y+1) = 0 \\ 2y + x^2 = 0 \end{cases} \implies \begin{cases} x = 0 \\ y = 0 \end{cases} \text{ or } \begin{cases} x = \pm\sqrt{2} \\ y = -1 \end{cases}.$$

Among these only $(0,0) \in D$ and $f(0,0) = 4$.

Now for boundary, four parts... whatever.

Remark

The reason why we need to check boundaries is because sometimes the critical points of the interior may not end up being the global maximum / minimum of the function over the domain. Consider

$$f(x,y) = |x^2 + y^2 - 1| \text{ over } D = \{(x,y) \mid x^2 + y^2 = 2\}.$$

The origin is the critical point with largest $f(x,y)$ in the interior but it is definitely not the global maximum — they appear on the boundary of D. On the other hand, just because

$$f(x,y) = x^2 + y^2$$

has no maximum in the interior doesn't mean it attains no global maximum on D: in fact they take place on the boundary of D.

10 Mon 10/12 11.8 Lagrange Multipliers, see CH12 Notes