In this class we have learned several concepts about multi-variable differentiation. Below is a summary of their definitions and relations to each other. For convenience we will only consider functions in two variables $x, y$. But they generalize to functions in three (or more) variables. (Another technicality: we will assume the domain of our function always includes a small disk containing the relevant point of differentiation.)

**Definition 0.1** (Partial derivatives). The partial derivative of $f$ with respect to $x$ at the point $P = (a, b)$ is defined as

$$
\partial_x f(P) \overset{def}{=} \frac{d}{dx} f(x, b) = \lim_{x \to a} \frac{f(x, b) - f(a, b)}{x - a}
$$

provided this limit exists. Similarly, the partial derivative of $f$ with respect to $y$ at the point $P = (a, b)$ is defined as

$$
\partial_y f(P) \overset{def}{=} \frac{d}{dy} f(a, y) = \lim_{y \to b} \frac{f(a, y) - f(a, b)}{y - b}
$$

provided this limit exists.

In words, in order to obtain $\partial_x f$, we fix $y = b$ and get a single-variable function $g(x) = f(x, b)$ and then differentiate $g$ according to single-variable calculus definition. Geometrically, we restrict movement in the domain to the horizontal line $y = b$. (The case for $\partial_y f$ is analogous.)

**Definition 0.2** (Gradient vector). If the partial derivatives of $f$ exist at $P$, the gradient vector of $f$ at $P$ is defined as

$$
\nabla f(P) \overset{def}{=} \langle \partial_x f(P), \partial_y f(P) \rangle.
$$

As we will see, this vector (which is a bit artificial at this moment) is only useful when more conditions are imposed on $f$.

**Definition 0.3** (Directional derivatives). The directional derivative of $f$ in the direction $u = \langle h, k \rangle$ at the point $P = (a, b)$ is

$$
D_uf(P) \overset{def}{=} \frac{d}{dt} f(P + tu) = \lim_{t \to 0} \frac{f(a + th, b + tk) - f(a, b)}{t}
$$

provided this limit exists.
In comparison to partial derivatives, we now restrict movement in the domain to the line $P + tu$. Directional derivatives include partial derivatives as special cases: when $u = (1, 0)$, this is the horizontal line $y = b$, and we get $\partial_x f$; when $u = (0, 1)$, this is the vertical line $x = a$, and we get $\partial_y f$.

The term “direction” requires some clarification: in general, the “direction vector” $u$ does not have to be a unit vector. It is better to think of it as “velocity vector”, the direction AND speed at which you move in the domain. Then we can think of directional derivative $D_u f(P)$ as the rate of change of the output $f(x, y)$ per unit of time when the input $(x, y)$ is moving at velocity $u$. If $u$ is a unit vector, this simplifies as the rate of change of output per unit of movement of input in the direction $u$.

This intuition is bolstered by the following fact: if $k$ is a real number, $u$ is a vector, then

$$D_{ku} f(P) = kD_u f(P).$$

For example, if the input is moving at twice the speed, then the output changes at twice the rate, and if the input is moving in the opposite direction, then the output also changes its direction of change (i.e. positive rate becomes negative rate, negative rate becomes positive rate).

What if instead of multiplying the direction vectors by a number, we add two direction vectors? One might expect the directional derivatives should also add, i.e. if $u_1, u_2$ are vectors, then

$$D_{u_1 + u_2} f(P) = D_{u_1} f(P) + D_{u_2} f(P).$$

This seems a reasonable guess: if the input is moving at the combined velocity $u_1 + u_2$, then the output should also change at the combined rate $D_{u_1} f(P) + D_{u_2} f(P)$.

**If this is true**, then the direction derivative is a **linear** function in the direction vector $u$, i.e. for any real numbers $k_1, k_2$ and vectors $u_1, u_2$, it holds that

$$D_{k_1 u_1 + k_2 u_2} f(P) = k_1 D_{u_1} f(P) + k_2 D_{u_2} f(P)$$

In particular, we can represent every vector $u = \langle h, k \rangle$ as $h\langle 1, 0 \rangle + k\langle 0, 1 \rangle$, so we get a formula for directional derivative

$$D_u f(P) = h\partial_x f(P) + k\partial_y f(P)$$

$$= \langle \partial_x f(P), \partial_y f(P) \rangle \cdot \langle h, k \rangle$$

$$= \nabla f(P) \cdot u.$$

(1)

just in terms of the partial derivatives!
However, this is not true in general. There are two problems that can happen. First, directional derivatives might not exist at a point when partial derivatives exist at that point.

**Example 0.4 (Partials but no directionals).** (Problem 7, Section 1 in final worksheet) Define $f(x, y) = \min(|x|, |y|)$.

1. Show that $f$ has partial derivatives at $(0, 0)$.
2. Show that $f$ has no directional derivatives in any direction other than the $x$- and $y$- directions.

Second, even when all directional derivatives exist at a point, the formula above might fail.

**Example 0.5 (Directionals not given by formula (1)).** Define $f(x, y) = \frac{x^2 y}{x^2 + y^2}$ for $(x, y) \neq 0$ and $f(0, 0) = 0$.

1. Show that $f$ is continuous at $(0, 0)$.
2. Show that $D_v f(0, 0)$ exists for every vector $v \neq 0$ by computing it.
3. Give example of $v \neq 0$ such that $D_v f(0, 0) \neq (f_x(0, 0), f_y(0, 0)) \cdot v$.

In order to get the formula (1), we need more condition than the existence of all directional derivatives. Let us now introduce a stronger notion of differentiability, which we will simply call “differentiable” because it is the most natural generalization of differentiation to the multi-variable situation.

**Definition 0.6 (Differentiability).** We say $f$ is **differentiable** at $P = (a, b)$ if

\[ f(x, y) = L(x, y) + E(x, y) \]

where $L(x, y)$ is a linear function (i.e. of the form $Ax + By + C$) and $E(x, y)$ satisfies

\[ \lim_{(x,y) \to P} \frac{|E(x, y)|}{|(x - a, y - b)|} = 0. \]

The linear function $L(x, y)$ is called the linear approximation of $f$ at $P$. The condition on $E(x, y)$ is often abbreviated as $E(x, y) = o(|(x - a, y - b)|)$; this is the "little oh" notation.

The key insight of this definition is that a function $f$ is differentiable at $P$ if it is like a linear function near $P$, in the sense that the output error decreases faster than the input distance to $P$.

This basic property, in turn, allows us to compute $L$: since the function $f$ and $L$ are basically the same near $P$, their value at $P$ must agree and
their partial derivatives at \( P \) must also agree, and a linear function is completely determined by these three data. Moreover, the formula (1) now holds. This formula critically depends upon the linearity of \( D_u f(P) \) in the direction vector \( u \). This assumption is now supplied by the linearity of \( L \). In summary, we have

**Theorem 0.7 (Gradient of a differentiable function).** If \( f \) is differentiable at \( P = (a, b) \), then

1. The linear approximation at \( P \) is given by
   \[
   L(x, y) = f(P) + \nabla f(P) \cdot (x - a, y - b).
   \]
2. \( f \) is continuous at \( P \).
3. \( f \) has all directional derivatives at \( P \) and
   \[
   D_u f(P) = \nabla f(P) \cdot u.
   \]

For this reason, if a function \( f \) is differentiable at \( P \), we call the gradient vector \( \nabla f(P) \) the (total) derivative of \( f \) at \( P \), and it is the unique vector that satisfies

\[
f(x, y) = f(P) + \nabla f(P) \cdot (x - a, y - b) + E(x, y)
\]
where \( E(x, y) = o(|(x - a, y - b)|) \) (for the meaning of this “little oh” notation, see Definition 0.6). Equivalently, it is the unique vector that satisfies

\[
\lim_{(x, y) \to P} \frac{|f(x, y) - f(P) - \nabla f(P) \cdot (x - a, y - b)|}{|(x - a, y - b)|} = 0
\]

If we recall the definition of derivative in single-variable calculus, the last equation suggests that the gradient vector/total derivative \( \nabla f \) is the multi-variable analogue of the single-variable derivative \( f' \).

Now we have a sufficient condition for formula (1): if \( f \) is differentiable at \( P \), then formula (1) holds. However, the reverse implication is false: a function \( f \) satisfying formula (1) can still fail to be differentiable. In other words, the condition of differentiability is strictly stronger than formula (1).

**Example 0.8 (Directionals given by formula (1) but not differentiable).** Define \( f(x, y) = \frac{x^3 y}{x^4 + y^2} \) for \( (x, y) \neq 0 \) and \( f(0, 0) = 0 \).

(1) Show that \( f \) is continuous at \((0, 0)\). (Hint: make the substitution \( z = x^2 \).)

(2) Show that \( D_v f(0, 0) = 0 \) for every vector \( v \neq 0 \). In particular, this shows that \( D_v f(0, 0) = (f_z(0, 0), f_y(0, 0)) \cdot v \) for every vector \( v \neq 0 \).
(3) Show that $f$ is not differentiable at $(0,0)$ by using the Linear Approximation definition. (Hint: approach $(0,0)$ along the parabola $y = x^2$.)

The next question is how can we know when a function $f$ is differentiable? We can check the definition, but this can be tedious and there is a simple criterion that suffices in many situations.

**Theorem 0.9** (Continuous partials implies differentiable). If all partial derivatives of $f$ are **continuous** at $P$, then $f$ is differentiable at $P$.

**Remark 0.10.** We are implicitly also requiring that all partial derivatives of $f$ exist in a small disk containing $P$, because it doesn't make sense to say something is continuous at $P$ unless it is defined on a small disk containing $P$.

The assumption of continuity of ALL partials can be in fact relaxed a little:

**Theorem 0.11** (slightly extended version of Theorem 0.9). If all partial derivatives of $f$ exist at $P$, and all but one partial derivatives are **continuous** at $P$, then $f$ is differentiable at $P$.

For example, if $\partial_x f(P)$ exists at $P$ and $\partial_y f$ is continuous at $P$, then $f$ is differentiable at $P$.

**Remark 0.12.** This assumption may seem strange but it actually naturally generalizes the case of dimension 1 (single-variable calculus). In dimension 1, there is only one “partial derivative” and if it exists at a point then the function is differentiable at that point by definition. It makes sense that the extra requirement of continuity is only needed for the partial derivatives in the extra dimensions.

Sometimes computing partial derivatives is also very tedious. If a function is a “combination” of functions that are already known to be differentiable, then it is also differentiable. More precisely,

**Theorem 0.13.** All statements below are at a point $P$

1. Sums/products: if $f, g$ are differentiable, so is $f + g, fg$.
2. Quotients: if $f, g$ are differentiable and $g(P) \neq 0$, then so is $f/g$.
3. Composition: if $f(x, y)$ is differentiable, and $g(z)$ is differentiable (in the single-variable sense), then so is $g(f(x, y))$.
4. Composition: if $f(x, y)$ is differentiable, and $x(t), y(t)$ are differentiable (in the single-variable sense), then so is $f(x(t), g(t))$. 
Combining these two theorems, we can easily deduce a variety of functions are differentiable, e.g. polynomials, exponentials, sines and cosines; sums, products, quotients of these (be careful with denominator equal to zero); various compositions of these, etc.

The next example shows the criterion of Theorem 0.9 (and its slightly extended version, Theorem 0.11) does not capture all differentiable functions: there exist differentiable functions whose partial derivatives are all discontinuous.

Example 0.14 (Differentiable but partials all discontinuous). (Problem 10, Section 1 in final worksheet) Define \( f(x, y) = \sin \left( \frac{1}{x^2 + y^2} \right) (x^2 + y^2) \) if \((x, y) \neq (0, 0)\) and \( f(0, 0) = 0 \) if \((x, y) = (0, 0)\).

(1) Show \( f \) is differentiable at \((0, 0)\) using the Linear Approximation definition.

(2) Show both partial derivatives of \( f \) are discontinuous at \((0, 0)\).

Summary of the discussion above:

The following conditions are strictly decreasing in strength:
(1) \( f \) has continuous partial derivatives at \( P \).
(2) \( f \) is differentiable at \( P \).
(3) \( f \) is continuous at \( P \), has directional derivatives at \( P \), and \( D_u f(P) = \langle \partial_x f(P), \partial_y f(P) \rangle \cdot u \).
(4) \( f \) is continuous at \( P \) and has directional derivatives at \( P \).
(5) \( f \) is continuous at \( P \) and has partial derivatives at \( P \).

i.e. (1) implies (2) implies (3) implies (4) implies (5), but none of the higher-indexed condition implies any lower-indexed condition.

1. Solutions to Examples

Example 0.4
(1) \( f(x, 0) = \min(|x|, 0) = 0 \), so \( \partial_x f(0, 0) = 0 \). Similarly, \( \partial_y f(0, 0) = 0 \).

(2) Take slope \( k \neq 0 \). On the line \( y = kx \): if \( |k| > 1 \), \( f(x, y) = \min(|x|, |k||x|) = |x| \), clearly not differentiable at 0; if \( |k| \leq 1, f(x, y) = \min(|x|, |k||x|) = |k||x| \), again clearly not differentiable at 0.

Example 0.5
(1) \( f \) is continuous at \((0, 0)\): use polar coordinates and squeeze theorem.
(2) \( D_v f(0, 0) = \frac{h^2k}{h^2 + k^2} \) for \( v = \langle h, k \rangle \neq 0 \).

(3) From the above we calculate \( f_x(0, 0) = f_y(0, 0) = 0 \). But clearly \( D_v f(0, 0) \neq 0 \) for most direction vectors \( v \), e.g. \( v = \langle 1, 1 \rangle \).

This means \( D_v f(0, 0) = \langle f_x(0, 0), f_y(0, 0) \rangle \cdot v \) does NOT hold in general.

Example 0.8

(1) \( f \) is continuous at \((0, 0)\): make the substitution \( z = x^2 \), use polar coordinates for \((z, y)\), and use squeeze theorem.

(2) For \( v = \langle h, k \rangle \neq 0 \), calculating using the definition of directional derivatives gives that \( D_v f(0, 0) = \lim_{t \to 0} \frac{t^2h^3k}{t^2h^4 + k^2} \). If \( k \neq 0 \), then this expression goes to zero as \( t \) goes to zero. If \( k = 0 \) (then \( h \) cannot be zero), this expression is already zero and so the limit equals zero as well.

(3) Near the point \((0, 0)\), the Linear Approximation term is zero because the function and its partial derivatives are both zero at \((0, 0)\). So the Error term equals the function \( f \) itself. This means we need to show that the limit of the expression

\[
\frac{E(x, y)}{|(x, y) - (0, 0)|} = \frac{f(x, y)}{\sqrt{x^2 + y^2}}
\]

\[
= \frac{x^3y}{(x^4 + y^2)\sqrt{x^2 + y^2}}
\]

does NOT equal zero as \((x, y) \to (0, 0)\). It is enough to show the limit does not exist along some path. Let’s take the path to be parabola \( y = x^2 \). Then the expression becomes

\[
\frac{x^3x^2}{(x^4 + x^4)\sqrt{x^2 + x^4}} = \frac{x^5}{2x^4|x|\sqrt{1 + x^2}}
\]

\[
= \frac{x/|x|}{2\sqrt{1 + x^2}}
\]

As \( x \to 0 \), this limit does not exist (more precisely, it has a right limit and left limit but they do not agree).

Example 0.14
(1) First we compute the partial derivatives at \((0, 0)\). Since \(f(x, 0) = \sin(1/x^4)x^2\) and \(f(0, 0) = 0\),
\[
\partial_x f(0, 0) = \lim_{x \to 0} \frac{f(x, 0) - f(0, 0)}{x - 0} = \lim_{x \to 0} \sin(1/x^4)x = 0
\]
by Squeeze Theorem. Similarly, \(\partial_y f(0, 0) = 0\). Thus, the linear approximation should be
\[
L(x, y) = f(0, 0) + \langle 0, 0 \rangle \cdot \langle x - 0, y - 0 \rangle = 0,
\]
so the error is
\[
E(x, y) = f(x, y) - L(x, y) = f(x, y).
\]
We need to show this error is going to zero faster than \(|(x - 0, y - 0)| = |(x, y)|\), i.e.
\[
\lim_{(x, y) \to (0, 0)} \frac{|f(x, y)|}{|(x, y)|} = 0.
\]
Plugging in the definition of \(f\), the left hand side equals
\[
\lim_{(x, y) \to (0, 0)} \frac{\sin\left(\frac{1}{x^4+y^4}\right)(x^2 + y^2)}{(x^2 + y^2)^{1/2}} = \lim_{(x, y) \to (0, 0)} \sin\left(\frac{1}{x^4+y^4}\right)(x^2+y^2)^{1/2} = 0
\]
by Squeeze Theorem.

(2) Let’s compute \(\partial_x f\) at an arbitrary point, not equal to \((0, 0)\).
Using the chain rule, the result is
\[
\partial_x f(x, y) = \cos\left(\frac{1}{x^4+y^4}\right) \cdot \frac{-4x^3}{(x^4+y^4)^2}(x^2 + y^2) + \sin\left(\frac{1}{x^4+y^4}\right) \cdot 2x
\]
The expression is quite messy. Let’s first take a look at the second term. By Squeeze Theorem, it’s going to zero as \((x, y) \to (0, 0)\).

Therefore the discontinuity must come from the first term. The first term is a product of a cosine term and a quotient of polynomials. The cosine term will oscillate infinitely many times between \(-1\) and 1. If we look at the exponent of polynomials in the numerator and the denominator, we see the denominator has exponent 8 and the numerator has exponent 5, so overall this will behave like exponent \(-3\), which will go to infinity as \((x, y) \to (0, 0)\). Overall, this term will fail to converge.

More precisely, let’s restrict to \(y = 0\). Then the expression simplifies as
\[
\partial_x f(x, 0) = \cos(1/x^4) \cdot \frac{-4x^3}{x^8}x^2 + \sin(1/x^4) \cdot 2x
\]
and as \(x \to 0\), the second terms goes to zero, but the first term will oscillate with larger and larger magnitudes.
2. *Proof of Theorem 0.9*

We will prove Theorem 0.11, which includes the conclusion of Theorem 0.9. Without loss of generality, we consider the case when \( \partial_x f(P) \) exists at \( P \) and \( \partial_y f \) is continuous at \( P \).

Let \( P = (a, b) \). We want to show that

\[
f(a + \Delta x, b + \Delta y) - f(a, b) = \partial_x f(a, b) \Delta x + \partial_y f(a, b) \Delta y + E(\Delta x, \Delta y)
\]

where \( E(\Delta x, \Delta y) \) goes to zero faster than \( (\Delta x, \Delta y) \).

The left hand side is the change in \( f \) when we move from \((a, b)\) to \((a + \Delta x, b + \Delta y)\). This is a movement in both x- and y-directions. We can split this change into two steps: in the first step, the change in \( f \) when we move in x-direction, from \((a, b)\) to \((a + \Delta x, b)\); in the second step, the change in \( f \) when we move in y-direction, from \((a + \Delta x, b + \Delta y)\).

\[
f(a+\Delta x, b+\Delta y) - f(a, b) = [f(a+\Delta x, b) - f(a, b)] + [f(a+\Delta x, b+\Delta y) - f(a+\Delta x, b)]
\]

Let’s first look at the first term on the right hand side of this equation. By definition of partial derivative, we know

\[
\lim_{\Delta x \to 0} \frac{f(a+\Delta x, b) - f(a, b)}{\Delta x} = \partial_x f(a, b)
\]

Equivalently,

\[
f(a + \Delta x, b) - f(a, b) = \partial_x f(a, b) \Delta x + E_1(\Delta x)
\]

where \( E_1(\Delta x) \) goes to zero faster than \( \Delta x \) goes to zero.

Let’s now look at the second term on the right hand side of equation (2). We could try to do the same thing, but the trouble is that the error function itself is no longer a fixed function but depends on where we are at after the first step. Instead, we will apply the Mean Value Theorem to partial differentiation in y:

\[
f(a + \Delta x, b + \Delta y) - f(a + \Delta x, b) = \partial_y f(a + \Delta x, b + \theta) \Delta y
\]

where \( \theta \) is some number between 0 and \( \Delta y \) (its exact value will depend on \( \Delta x, \Delta y \)). This is quite close to what we want, except the y-partial is not taken at \((a, b)\) but at \((a + \Delta x, b + \theta)\). However, as \((\Delta x, \Delta y)\) goes to zero, \((a + \Delta x, b + \theta)\) will go to \((a, b)\). Since \( \partial_y f \) is continuous at \((a, b)\), this means \( \partial_y f(a + \Delta x, b + \theta) \) will go to \( \partial_y f(a, b) \). Let’s denote

\[
E_2(\Delta x, \Delta y) = \partial_y f(a + \Delta x, b + \theta) - \partial_y f(a, b)
\]

Then we know \( E_2 \) goes to zero as \((\Delta x, \Delta y)\) goes to zero, and

\[
f(a + \Delta x, b + \Delta y) - f(a + \Delta x, b) = (\partial_y f(a, b) + E_2(\Delta x, \Delta y)) \Delta y
\]

\[
= \partial_y f(a, b) \Delta y + E_2(\Delta x, \Delta y) \Delta y
\]
Now putting everything together, we have

\[
f(a + \Delta x, b + \Delta y) - f(a, b) = \partial_x f(a, b) \Delta x + \partial_y f(a, b) \Delta y + E_1(\Delta x) + E_2(\Delta x, \Delta y) \Delta y
\]

\[
= \partial_x f(a, b) \Delta x + \partial_y f(a, b) \Delta y + E_1(\Delta x) + E_2(\Delta x, \Delta y) \Delta y
\]

We just need to verify that \(E_1(\Delta x) + E_2(\Delta x, \Delta y) \Delta y\) goes to zero faster than \(|(\Delta x, \Delta y)|\). This is clear for the first term \(E_1(\Delta x)\) because it goes to zero faster than \(\Delta x\) goes to zero. For the second term, we note that

\[
\lim_{(\Delta x, \Delta y) \to (0,0)} \frac{E_2(\Delta x, \Delta y) \Delta y}{|(\Delta x, \Delta y)|} = 0
\]

by Squeeze Theorem (\(E_2\) goes to zero and the \(\frac{\Delta y}{|(\Delta x, \Delta y)|}\) is bounded between \(-1\) and \(1\)).