Course notes for multivariable calculus

by Alec Johnson

spring–fall, 2009

1 Course overview

The purpose of this course is to generalize math 221 (single-variable calculus) to multiple variables.

Calculus is the study of things that are **smooth**. Smooth means *locally flat*. The study of things that are flat is called **Linear Algebra**. Calculus and linear algebra are the two foundational subjects for science, engineering, and most of the rest of mathematics (e.g. differential equations, probability, statistics).

This course covers Chapters 13 through 16 of the text:

1. **Chapter 13** covers the calculus of **curves**, i.e. smooth functions from $\mathbb{R}$ to $\mathbb{R}^2$ or $\mathbb{R}^3$. We represent a typical such function as $\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k} = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$. We think of t as time and $\mathbf{r}(t)$ as position in space.

2. **Chapter 14** covers the *differential* calculus of functions from $\mathbb{R}^n$ to $\mathbb{R}$. We represent a typical function from $\mathbb{R}^2$ to $\mathbb{R}$ by $f(x, y)$. The graph of such a function f is a **surface**. We will use *partial derivatives* (where we differentiate f with respect to one variable while holding the other variable(s) constant) to find the *tangent plane* at a given point on a surface.

3. **Chapter 15** covers the *integral* calculus of functions from $\mathbb{R}^n$ to $\mathbb{R}$. For example, we might want to find the volume under a surface $f(x, y)$ over a region in the x–y plane, or the total mass within a 3-dimensional region.

4. **Chapter 16** covers the differential and integral calculus of **vector fields**, i.e. smooth functions from $\mathbb{R}^n$ to $\mathbb{R}^n$. We represent a typical such function by $\mathbf{F}(\mathbf{r})$. (Usually we just write $\mathbf{F}$ with the understanding that it is a function of space $\mathbf{r}$.)

2 Notes on Chapter 12

2.1 Vectors

2.1.1 Points versus Vectors

In this course we will study points and vectors in 2- and 3-dimensional space. It is important to understand the distinction between a point and a vector.

A **point** is a position in space. Mathematically we represent a point using its coordinates in a coordinate system. The text usually uses capital letters to represent points. The text denotes a point in a Cartesian coordinate system using three coordinates in parentheses. To name the coordinate variables, we usually use either numerical subscripts or successive letters of the alphabet. For example: $P = (x, y, z)$, or $U = (u_1, u_2, u_3)$.

A **vector** is an "arrow": it has a magnitude and a direction. *Two vectors are the same if they have the same length and direction, even if their tails are anchored at different base points.* To represent a vector in a Cartesian coordinate system we place its tail at the origin and record the position of its head. The text tends to use bold lower case letters to stand for vectors; it represents a vector in a Cartesian coordinate system using three coordinates in angle brackets. For example: $\mathbf{r} = \langle x, y, z\rangle$, or $\mathbf{u} = \langle u_1, u_2, u_3\rangle$.

You can perform arithmetic on vectors and points as follows. It does not make sense to add two points. But you can take the difference of two points P and Q, and the result is a vector $\mathbf{v} = P - Q$. This vector represents the displacement from P to Q. If you add a vector to a point, you get another point, and if you add a vector to a vector you get another vector. For example: Let $\mathbf{v} = P - Q$. Let $\mathbf{u} = Q - R$. Let $\mathbf{w} = \mathbf{u} + \mathbf{v}$. Then $P = Q + \mathbf{v}$ and $Q = R + \mathbf{u}$, so $P = R + \mathbf{u} + \mathbf{v} = R + \mathbf{w}$.

In these notes we identify every point P with the vector $\mathbf{p}$ that points from the origin O to P. This allows us to blur the distinction between vectors and points.

2.1.2 Multiplication by a Scalar

To multiply a vector $\mathbf{u}$ by a scalar t, multiply each component by the scalar:

$$t\mathbf{u} = t\langle u_1, u_2, u_3\rangle := \langle tu_1, tu_2, tu_3\rangle.$$

(Note that "$A = B$" simply means that A and B are equal, whereas when we write "$A := B$" we are saying that A is *defined* to be B.) This rescales the length of $\mathbf{u}$ by a factor of t. If t is positive the direction remains the same; if t is negative the direction is reversed.

2.1.3 Dot Product

The dot product takes two vectors and gives you a scalar (i.e. a number). It is also called the **scalar product**. The dot product has an algebraic definition and a geometric definition. Algebraically the dot product of two vectors is the sum of the products of the corresponding components:

$$\mathbf{u} \cdot \mathbf{v} := u_1v_1 + u_2v_2 + u_3v_3.$$

From this definition you can easily show that the dot product obeys distributive, commutative, and associative laws:

property	**identity**
commutativity	$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
distributivity	$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
scalar associativity	$t(\mathbf{u} \cdot \mathbf{v}) = (t\mathbf{u}) \cdot \mathbf{v}$

You probably have seen "$\cdot$" used to denote multiplication by a scalar. This should not cause confusion, since the dot product of two scalars *is* their scalar product.

2.1.4 Norms

The length of a vector is called its **magnitude** or **norm**. The magnitude of a vector $\mathbf{v}$ is denoted $|\mathbf{v}|$, using absolute value symbols. This should not cause confusion, because for one-dimensional vectors the norm *is* the absolute value. But to be extra clear, we often use two bars for the norm and one bar for the absolute value. For example:

$$\|t\mathbf{v}\| = |t| \cdot \|\mathbf{v}\|.$$

You can apply the Pythagorean theorem to a couple right triangles to show that the square of the length of a vector is the sum of the squares of its components:

$$\|\mathbf{u}\|^2 = \mathbf{u} \cdot \mathbf{u} = u_1^2 + u_2^2 + u_3^2,$$

where $\|\mathbf{u}\|$ denotes the length of the vector $\mathbf{u}$.

2.1.5 Unit direction vectors

If we scale a vector by the reciprocal of its magnitude we will get a vector of length 1 called the **unit direction vector**. We often denote a direction vector by putting a hat over it. So we will write:

$$\boxed{\widehat{\mathbf{u}} := \frac{\mathbf{u}}{\|\mathbf{u}\|}}.$$

Observe that indeed $\|\widehat{\mathbf{u}}\|^2 = \widehat{\mathbf{u}} \cdot \widehat{\mathbf{u}} = \frac{\mathbf{u} \cdot \mathbf{u}}{\|\mathbf{u}\|^2} = 1$.

Three special unit vectors are the unit vectors along the principle axes (in the positive direction). These vectors are called the **standard basis** vectors. Different people give them different names:

$$\begin{aligned}
\widehat{\mathbf{e}}_1 &:= \widehat{\mathbf{x}} := \widehat{\mathbf{i}} := \mathbf{i} := \langle 1, 0, 0\rangle, \\
\widehat{\mathbf{e}}_2 &:= \widehat{\mathbf{y}} := \widehat{\mathbf{j}} := \mathbf{j} := \langle 0, 1, 0\rangle, \\
\widehat{\mathbf{e}}_3 &:= \widehat{\mathbf{z}} := \widehat{\mathbf{k}} := \mathbf{k} := \langle 0, 0, 1\rangle.
\end{aligned}$$

The book uses $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$.

2.1.6 Geometric definition of dot product (Law of Cosines)

If you anchor the tails of two vectors $\mathbf{u}$ and $\mathbf{v}$ at the origin, they span an angle θ and a triangle with sides of length $\|\mathbf{u}\|$, $\|\mathbf{v}\|$, and $\|\mathbf{v} - \mathbf{u}\|$. If you apply the law of cosines to the side of this triangle and simplify, you get the law of cosines for vectors, also known as the geometric definition of the dot product:

$$\boxed{\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \cdot \|\mathbf{v}\| \cos\theta}.$$

This says that the dot product of two vectors is the product of the lengths times the cosine of the angle between them. The geometric definition reveals the most important property of the dot product:

> *Two nonzero vectors are perpendicular if and only if their dot product is zero.*

2.1.7 Orthogonal decomposition and projection.

Given two vectors $\mathbf{u}$ and $\mathbf{v}$, we can use the dot product to write $\mathbf{v}$ as the sum of a vector $\mathbf{v}_{\|} = t\mathbf{u}$ parallel to $\mathbf{u}$ and a vector $\mathbf{v}_{\perp}$ perpendicular to $\mathbf{u}$:

$$\mathbf{v} = t\mathbf{u} + \mathbf{v}_{\perp}.$$

To find t dot this equation with $\mathbf{u}$ and solve for t. Since $\mathbf{v}_{\perp} \cdot \mathbf{u} = 0$, $t = \frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$. The vector $\mathbf{v}_{\|}$ is called the **projection** onto $\mathbf{u}$ of $\mathbf{v}$, which the book denotes as $\mathrm{pr}_{\mathbf{u}}\mathbf{v}$. So:

$$\boxed{\mathrm{pr}_{\mathbf{u}}\mathbf{v} := \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u} = (\mathbf{v} \cdot \widehat{\mathbf{u}})\widehat{\mathbf{u}} = \mathrm{pr}_{\widehat{\mathbf{u}}}\mathbf{v}.}$$

2.2 Physical meaning and application of dot product

Recall that $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|(\|\mathbf{v}\| \cos\theta)$. Since $\|\mathbf{v}\| \cos\theta$ is the length of the projection of $\mathbf{v}$ onto $\mathbf{u}$, the geometric definition of the dot product says:

> *The dot product of* $\mathbf{u}$ *and* $\mathbf{v}$ *is the length of* $\mathbf{u}$ *times the length of the projection of* $\mathbf{v}$ *onto* $\mathbf{u}$.

An important application is the definition of work. If a force $\mathbf{F}$ is applied to move an object through a displacement $d\mathbf{x}$, the amount of work dW is:

$$dW = \mathbf{F} \cdot d\mathbf{x},$$

i.e.,

> *The work performed when a force* $\mathbf{F}$ *is applied over a displacement* $d\mathbf{x}$ *is the magnitude of the displacement times the magnitude of the component of the force in the direction of the displacement, which is the same as the magnitude of the force times the magnitude of the component of the displacement in the direction of the force.*

2.2.1 Cross Product

In general, the cross product takes two vectors and gives a vector perpendicular to both of them.

Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors. Geometrically the cross product $\mathbf{w} = \mathbf{u} \times \mathbf{v}$ is defined to satisfy three properties:

1. $\mathbf{w}$ is perpendicular to $\mathbf{u}$ and $\mathbf{v}$; more precisely, $\mathbf{w} \cdot \mathbf{u} = 0$ and $\mathbf{w} \cdot \mathbf{v} = 0$,
2. the length of $\mathbf{w}$ is the area of the parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$; i.e., $\|\mathbf{w}\| = \|\mathbf{u}\| \cdot \|\mathbf{v}\| \sin\theta$, where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$, and
3. the ordered triple $\mathbf{u}, \mathbf{v}, \mathbf{w}$ has the same (conventionally right-handed) orientation as the standard basis vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$.

Algebraically the cross product $\mathbf{w} = \mathbf{u} \times \mathbf{v}$ is defined by

$$\begin{aligned} w_1 &:= u_2 v_3 - u_3 v_2, \\ w_2 &:= u_3 v_1 - u_1 v_3, \\ w_3 &:= u_1 v_2 - u_2 v_1. \end{aligned}$$

(You only need to remember the formula for one of the components. To get the other two formulas, you can just cycle the components using mod-3 cyclical arithmetic, where 4=1, 5=2, 6=3, etc.) You can easily verify that $\mathbf{w} \cdot \mathbf{u} = 0$ and $\mathbf{w} \cdot \mathbf{v} = 0$.

The cross product has the following properties:

property	**identity**
*anti*commutativity	$\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$
distributivity	$\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$
scalar associativity	$t(\mathbf{u} \times \mathbf{v}) = (t\mathbf{u}) \times \mathbf{v}$

Note that it is *not* true in general that $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times (\mathbf{v} \times \mathbf{w})$, i.e., the cross product is *not* associative. (Try it out with the standard basis vectors.)

2.3 Lines

A line is determined by a point $\mathbf{r}_0$ on the line and a vector $\mathbf{u}$ in the direction of the line. A generic point $\mathbf{r}$ is on the line if $\mathbf{r} - \mathbf{r}_0$ is parallel to $\mathbf{u}$, i.e., if there exists a scalar t such that $\mathbf{r} - \mathbf{r}_0 = t\mathbf{u}$. Solving for $\mathbf{r}$ gives an equation for the line in terms of the parameter t:

$$\mathbf{r}(t) = \mathbf{r}_0 + t\mathbf{u}.$$

Written out in components, this is:

$$\begin{aligned} x &= x_0 + tu_1, \\ y &= y_0 + tu_2, \\ z &= z_0 + tu_3. \end{aligned}$$

To eliminate the parameter t and get a system of equations in x, y, and z we solve each equation for t:

$$t = \frac{x - x_0}{u_1} = \frac{y - y_0}{u_2} = \frac{z - z_0}{u_3}.$$

This is a system of two independent equations in three unknowns. The graph of each equation is a plane. Their intersection is our line. Note that this system is not uniquely determined, since we could rescale $\mathbf{u}$ or choose a different point $\mathbf{r}_0$ on the line.

2.4 Planes

A plane is determined by a point $\mathbf{r}_0$ on the plane and a vector $\mathbf{n} = (A, B, C)$ perpendicular to the plane. The condition for a generic point $\mathbf{r}$ to be on the plane is that the difference vector $\mathbf{r} - \mathbf{r}_0$ must lie in the plane, i.e., it must be perpendicular to $\mathbf{n}$:

$$\begin{aligned} (\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{n} = 0, &\text{ i.e.,} \\ \mathbf{r} \cdot \mathbf{n} - \mathbf{r}_0 \cdot \mathbf{n} = 0, &\text{ i.e.,} \\ \mathbf{r} \cdot \mathbf{n} = \mathbf{r}_0 \cdot \mathbf{n}&. \end{aligned}$$

This says that *any two vectors in the plane have the same dot product D with n*. Written out in components this reads

$$Ax + By + Cz = D.$$

Observe that you can read off the coefficients of x, y, and z to get the components of the normal vector.

2.5 Exercises

The following exercises should serve as a quick check on your understanding of the most important skills and concepts of chapter 12:

1. Write the vector $\mathbf{v} = \langle 2, -3, 4\rangle$ as the sum of vectors parallel and perpendicular to the vector $\mathbf{u} = \langle 3, -4, 12\rangle$. Check your answer.
2. Find parametric and symmetric equations for the line through the points $P = (-1, 3, -2), Q = (1, 2, 4)$. (Hint: the difference of two different points in a line is a vector in the direction of the line.) Check that both points are on the line.
3. Find the equation of the plane through the three points $P = (-1, 3, -2), Q = (1, 2, 4), R = (0, 4, 5)$. (Hint: find two nonparallel vectors that lie in the plane and take their cross product to get a vector perpendicular to the plane.) Check that all three points satisfy the equation of the plane.

2.6 Quadratics

By rotation, shifting, and rescaling of axes, every nondegenerate quadratic in three variables (i.e. any expression of the form $A_{11}x^2 + A_{22}y^2 + A_{33}z^2 + 2A_{12}xy + 2A_{13}xz + 2A_{23}yz + B_1x + B_2y + B_3z + F = 0$) can be put in one of the following forms:

Form	**Type of Quadric Surface**
$x^2 + y^2 + z^2 = 1$	ellipsoid
$x^2 + y^2 - z^2 = 1$	hyperboloid of one sheet
$-x^2 - y^2 + z^2 = 1$	hyperboloid of two sheets
$x^2 + y^2 - z = 0$	elliptic paraboloid
$x^2 - y^2 - z = 0$	hyperbolic paraboloid
$x^2 + y^2 - z^2 = 0$	cone
$x^2 + y^2 = 1$	elliptic cylinder
$x^2 - y^2 = 1$	hyperbolic cylinder
$x^2 + y = 0$	parabolic cylinder.

To understand the graphs of equations that involve the sum of two squares, recall that in cylindrical coordinates $r^2 = a^2 + b^2$ and graph z versus r. The hyperbolic paraboloid looks like a saddle. To see this it helps to look at slices. See section 12.6 and `http://en.wikipedia.org/wiki/Quadratic_surface` for more details.

3 Notes on Chapter 13

Definitions:

- A dot over a letter means the derivative with respect to t ("time").
- $\mathbf{r}(t)$ = position as a function of time
- $\mathbf{v} := \dot{\mathbf{r}}$ = velocity
- $\mathbf{a} := \dot{\mathbf{v}} = \ddot{\mathbf{r}}$ = acceleration
- s is distance along the curve (measured from some reference point.
- $\dot{s} := \frac{ds}{dt} = |\mathbf{v}|$ is the *speed.*
- $\hat{\mathbf{T}} := \frac{d\mathbf{r}}{ds} = \frac{d\mathbf{r}}{dt}\frac{dt}{ds} = \frac{\mathbf{v}}{|v|}$ is the unit tanget vector.
- $\vec{\kappa} := \frac{d\hat{\mathbf{T}}}{ds}$ is the curvature vector
- $\kappa = |\vec{\kappa}|$ is the curvature.
- $\hat{\mathbf{N}} := \frac{\vec{\kappa}}{|\vec{\kappa}|}$ is the unit normal vector (i.e. the direction of the curvature vector)

Quantities that are defined in terms of the arc length, such as the unit tangent and the curvature, do not depend on the speed of the parametrization, but only on its physical shape.

Definition 3.1 (Arc length)**.** Since speed is the time-derivative of arc length, arc length is the integral of speed:

$$L = \int_{s_0}^{s_1} ds = \int_{t_0}^{t_1} \frac{ds}{dt} dt$$

$$\boxed{L = \int_{t_0}^{t_1} |\mathbf{r}'(t)| dt}$$

Problem 3.1. *Given* $\mathbf{r}, \mathbf{v}, \mathbf{a}$*, find* $\hat{\mathbf{T}}, \vec{\kappa}$*.*

Solution.

$$\mathbf{v} = \hat{\mathbf{T}}\dot{s}.$$

So differentiating with respect to time,

$$\begin{aligned} \mathbf{a} &= \frac{d\hat{\mathbf{T}}}{ds}\frac{ds}{dt}\dot{s} + \hat{\mathbf{T}}\ddot{s} \\ &= \underbrace{\vec{\kappa}\dot{s}^2}_{\mathbf{a}_N} + \underbrace{\hat{\mathbf{T}}\ddot{s}}_{\mathbf{a}_T} \end{aligned}$$

By parametrizing a circle one can easily show that the curvature is the reciprocal of the radius R of the circle. So $a_N = \dot{s}^2|\kappa| = |v|^2/R$, a familiar formula from physics.

The tangential component of the acceleration is just its projection onto the velocity vector $\mathbf{v}$:

$$|\mathbf{a}_T| = \mathbf{a} \cdot \hat{\mathbf{T}}, \qquad \mathbf{a}_T = |\mathbf{a}_T|\hat{\mathbf{T}}.$$

The normal component is then

$$\mathbf{a}_N = \mathbf{a} - \mathbf{a}_T, \qquad |\mathbf{a}_N|^2 = |\mathbf{a}|^2 - |\mathbf{a}_T|^2,$$

The curvature is then

$$\vec{\kappa} = \frac{\mathbf{a}_N}{|v|^2}, \qquad |\vec{\kappa}| = \frac{|\mathbf{a}_N|}{|v|^2}.$$

There is a shortcut to find the magnitude of the curvature:

$$|\mathbf{a}_N| = |\mathbf{a}_N \times \hat{\mathbf{T}}| = |(\mathbf{a}_N + \mathbf{a}_T) \times \hat{\mathbf{T}}|,$$

so

$$|\vec{\kappa}| = \frac{|\mathbf{a} \times \mathbf{v}|}{|\mathbf{v}|^3}.$$

□

4 Notes on Chapter 14

This section deals with smooth real-valued functions of multiple variables, e.g. $f(x, y)$.

4.1 Partial derivatives

The partial derivative of a function $f(x, y, z)$ with respect to x is just the ordinary derivative with respect to x holding the other variables constant. It is the rate of change of f as you move along the x axis. It is denoted $\frac{\partial f}{\partial x}$, $D_x f$, or f_x.

Remark 4.1 (Notation for partial derivatives)**.** *The notation for partial derivatives is problematic. When you take a partial derivative, you must be very clear what function and what argument you are talking about.* $\frac{\partial f}{\partial x}$ *means "the partial derivative of* f *with respect to the formal argument named* x *holding the other arguments constant". The problem comes in when you take the derivative of an* expression*, where the function is not explicitly defined. Then you need to be clear "which arguments you are holding constant", i.e., which function you are talking about. Here is an example of ambiguity that arises when you are doing implicit partial differentiation: Does* $\frac{\partial f(x,y,z(x,y))}{\partial x}$ *mean*

$$\frac{\partial}{\partial x}\Big[(x, y) \mapsto F(x, y, z(x, y))\Big]\Big|_{(x,y)} =: F(x, y, z(x, y))_x$$

or

$$\frac{\partial}{\partial x}\Big[(x,y,z)\mapsto F(x,y,z)\Big]\Big|_{(x,y,z(x,y))} =: F_x(x,y,z(x,y))?$$

The problem is that our notation for function evaluation is ambiguous and uses the letter x to refer to two different things: (1) the first argument of the function f and (2) the first coordinate of the point *where we are evaluating our partial derivative. One way to deal with this would be to rename the arguments of f as (u,v,w). Another way is to use the notation D_n, the partial derivative with respect to the nth argument.*

4.2 Linear approximation

A function is smooth (i.e. differentiable) if you can approximate it by a linear function.

Proposition 4.2 (Linear approximation)**.** *Near (x_0,y_0), $f(x,y)\approxeq L(x,y)$, where the linear approximation $L(x,y)$ is given by*

$$\begin{aligned} L(x,y) := f(x_0,y_0) &+ (x-x_0)\frac{\partial f}{\partial x}\Big|_{(x_0,y_0)} \\ &+ (y-y_0)\frac{\partial f}{\partial y}\Big|_{(x_0,y_0)} \end{aligned}$$

This should seem intuitively clear to you. It says that when you move from (x_0,y_0) to (x,y), the change in f is approximately the rate at which f changes as you move along the x axis times the change in x plus the rate at which f changes as you move along the y axis times the change in y.

The language of **differentials** is designed to make this clear. A small change in x and y is denoted by the "differentials" $dx := x-x_0$ and $dy := y-y_0$ and the resulting small change in f is approximated by the "differential" df.

Definition 4.3 (Differential)**.**

$$f(x,y)-f(x_0,y_0)\approx df, \quad \text{where}$$

$$\boxed{df := dx\frac{\partial f}{\partial x}\Big|_{(x_0,y_0)} + dy\frac{\partial f}{\partial y}\Big|_{(x_0,y_0)}}$$

Remark: $L(x,y) = f(x_0,y_0)+df$.

In my opinion, the best way to make sure that you understand a calculus relationship is to see what it tells you in the linear case. That is, if you want to understand why a rule is true, try it out on a linear function. So take a linear function $f(x,y) = Ax+By+C$ and plug it into the equations above. These approximate equalities should become exact equalities.

The formal definition of differentiability basically says that the error of the linear approximation is small:

Definition 4.4 (Derivative)**.** $f(\mathbf{r})$ is **differentiable** at $(\mathbf{r}_0)$ if there is a linear approximation $L(\mathbf{r}) = C + \mathbf{n}\cdot\mathbf{r} = C+Ax+By$. ($L$ is a linear approximation at $\mathbf{r}_0$ if as $\mathbf{r}$ goes to $\mathbf{r}_0$ the error $f(\mathbf{r})-L(\mathbf{r})$ goes to zero even faster (i.e., $\lim_{|\mathbf{r}-\mathbf{r}_0|\to 0}\frac{f(\mathbf{r})-L(\mathbf{r})}{|\mathbf{r}-\mathbf{r}_0|}=0$)).

4.3 Chain Rule

Proposition 4.5 (Chain rule for differentials)**.** *Given the functions $f(u,v)$, $u(t)$, and $v(t)$, taking the differential and applying the one-variable chain rule gives*

$$\begin{aligned} df &= \frac{\partial f}{\partial u}du \quad + \frac{\partial f}{\partial v}dv \\ &= \frac{\partial f}{\partial u}\frac{du}{dt}dt + \frac{\partial f}{\partial v}\frac{dv}{dt}dt. \end{aligned}$$

Again, to see why this is true, just try it out with linear functions. Dividing the chain rule for differentials by dt gives:

Proposition 4.6 (Chain rule (I))**.**

$$\boxed{\frac{df}{dt} = \frac{\partial f}{\partial u}\frac{du}{dt} + \frac{\partial f}{\partial v}\frac{dv}{dt}}.$$

Proposition 4.7 (Chain rule (II))**.** *In case $u(t,s)$ and $v(t,s)$, holding s constant and applying chain rule (I) gives*

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial u}\frac{\partial u}{\partial t} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial t}.$$

4.4 Directional derivatives and the gradient

Definition 4.8 (Directional derivative)**.** The derivative of $f(x,y)$ in the direction $\hat{\mathbf{u}}$ near the point

$\mathbf{r}_0 = (x_0, y_0)$ is just the rate of change of f as you travel through $\mathbf{r}_0$ along a line in the direction $\hat{\mathbf{u}}$, i.e., $\frac{df(\mathbf{r}(t)}{dt}$, where $\mathbf{r}(t) = \mathbf{r}_0 + t\hat{\mathbf{u}}$:

$$\begin{aligned} D_{\hat{\mathbf{u}}} f\Big|_{\mathbf{r}_0} &= \frac{df(\mathbf{r}_0 + t\hat{\mathbf{u}})}{dt} \\ &= \frac{dx}{dt}\frac{\partial f}{\partial x} + \frac{dy}{dt}\frac{\partial f}{\partial y} \\ &= u_1 \frac{\partial f}{\partial x} + u_2 \frac{\partial f}{\partial y} \\ &= \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \end{aligned}$$

In other words,

$$\boxed{D_{\hat{\mathbf{u}}} f = \hat{\mathbf{u}} \cdot \nabla f = |\nabla f| \cos(\theta)},$$

where $\nabla f := (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$ and θ is the angle between $\hat{\mathbf{u}}$ and the gradient vector ∇f. Thus *the directional derivative of f varies between $-|\nabla f|$ and $|\nabla f|$ and is maximized in the direction of ∇f.*

4.5 Implicit differentiation

Problem 4.1 (Implicit partial derivative)**.** *Suppose that near $\mathbf{r}_0 = (x_0, y_0, z_0)$ the function $f(x, y, z)$ is smooth and $\frac{\partial f}{\partial z} \neq 0$. Then the equation*

$$f(x, y, z) = 0 \tag{1}$$

implicitly defines a function $z(x, y)$ near $\mathbf{r}_0$. To find the partial derivatives of z, we can use the chain rule to differentiate $f(x, y, z(x, y))$, being very clear with notation. As a shortcut, take the differential of (1) *and get*

$$\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz = 0. \tag{2}$$

If we hold y constant, then $dy = 0$. Then

$$\boxed{\frac{\partial z}{\partial x} = \frac{-\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial z}}}. \tag{3}$$