1. Return to derivatives in one variable

Recall that if \( f \) is a real- or complex-valued function defined on an interval \((a, b) = \{ x \in \mathbb{R} : a < x < b \}\), and if \( x \in (a, b) \), then the function \( f \) is differentiable at \( x \) if and only if the limit

\[
\lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = f'(x)
\]  

exists. We want to rewrite this definition in a way that makes it easier to extend the concept of derivative to the case of functions of several variables. For \( h \neq 0 \), set

\[
E(x, h) \equiv h \left[ \frac{f(x + h) - f(x)}{h} - f'(x) \right].
\]  

Then we can rewrite (1.0.1) as

\[
f(x + h) = f(x) + f'(x)h + E(x, h)
\]  

and the function \( f \) is differentiable at the point \( x \) if and only if

\[
\lim_{h \to 0} \frac{|E(x, h)|}{|h|} = 0.
\]  

Now we think of the right hand side of equation (1.0.3) as a function of \( h \) which is the sum of a constant function plus a linear function of \( h \) plus an error term. The error term goes to zero as \( h \) goes to zero, and in fact, it is “second order small”. Equation (1.0.4) says that even after dividing by the quantity \( |h| \), the quotient still goes to zero.

Thus we can reinterpret the concept of differentiability as follows. The function \( f \) is differentiable at a point \( x \) if near \( x \) the function equals a constant function plus a linear function of the change plus an error which is second order small. The ‘derivative’ is thus a linear mapping from \( \mathbb{R} \) to \( \mathbb{R} \).

2. The case of several variables

We now want to consider real- or complex-valued functions of several variables, and in fact, we want to allow such functions to take values in spaces of several variables. Thus we want to consider functions

\[
F : \mathbb{R}^n \to \mathbb{R}^n \quad \text{or} \quad F : \mathbb{C}^n \to \mathbb{C}^m,
\]

even more generally, we may want to consider functions from one vector space \( V \) to another vector space \( W \). Very often, the functions we want to study are not defined on all of \( \mathbb{R}^n \) or all of \( V \), and so we need to discuss the kind of subsets on which it may be defined. Also, we will need to understand what we mean by limits in the case of several variables.

2.1. Open balls and open sets.

Recall that for points \( \mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \) or \( \mathbf{z} = (z_1, \ldots, z_n) \in \mathbb{C}^n \) we have introduced the notion of the length or norm

\[
||\mathbf{x}|| = \sqrt{x_1^2 + \cdots + x_n^2}, \quad ||\mathbf{z}|| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.
\]  

We also saw that this norm satisfies the triangle inequality:

\[
||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||, \quad ||\mathbf{z} + \mathbf{w}|| \leq ||\mathbf{z}|| + ||\mathbf{w}||.
\]  

From now on we will focus mainly on the case of \( \mathbb{R}^n \), although everything we do will have an analogue for \( \mathbb{C}^n \).
Definition 2.1.

(1) The \textbf{ball} centered at a point \(x \in \mathbb{R}^n\) of radius \(r\) is the set
\[
B(x, r) = \{y \in \mathbb{R}^n : ||x - y|| < r\}. \tag{2.1.3}
\]

(2) If \(S \subseteq \mathbb{R}^n\) is a subset, a point \(x \in S\) is an \textbf{interior point} if there exists \(r > 0\) so that \(B(x, r) \subseteq S\).

(3) A set \(S \subseteq \mathbb{R}^n\) is \textbf{open} if every point in \(S\) is an interior point.

(4) A set \(E \subseteq \mathbb{R}^n\) is \textbf{closed} if the complement \(E^c = \{x \in \mathbb{R}^n : x \notin E\}\) is an open set.

(5) If \(S \subseteq \mathbb{R}^n\) is a subset, a point \(x \in S\) is an \textbf{exterior point} if there exists \(r > 0\) so that \(B(x, r) \cap E = \emptyset\).

(6) The \textbf{boundary} of a set \(E\) is the set of points which are neither interior points of \(E\) nor exterior points of \(E\).

Lemma 2.2.

(a) If \(x \in \mathbb{R}^n\) and \(r > 0\), then every point of \(B(x, r)\) is an interior point, and hence \(B(x, r)\) is an open set.

(b) If \(y \in \mathbb{R}^n\), then \(y\) is an exterior point of \(B(x, r)\) if and only if \(||x - y|| > r\).

(c) The boundary of \(B(x, r)\) is the set \(S(x, r) = \{y \in \mathbb{R}^n : ||x - y|| = r\}\).

2.2. Limits and Continuity.

We will use \(||x||_n\) for the norm of a vector \(x \in \mathbb{R}^n\) and \(||y||_m\) for the norm of a vector \(y \in \mathbb{R}^m\). Let \(S \subseteq \mathbb{R}^n\), and let \(F : S \rightarrow \mathbb{R}^m\).

Definition 2.3. Let \(a \in \mathbb{R}^n\) and \(b \in \mathbb{R}^m\). Then
\[
\lim_{x \to a} F(x) = b
\]
means that for every \(\epsilon > 0\) there exists \(\delta > 0\) so that if \(x \in S\) and \(||x - a||_n < \delta\) then \(||F(x) - b||_m < \epsilon\). Equivalently, this means that
\[
(\forall \epsilon > 0) \ (\exists \delta > 0) \ (x \in B(a, \delta) \implies F(x) \in B(b, \epsilon))
\]
Note that it is not required that \(x\) be an element of \(S\).

Now suppose that \(F : S \rightarrow \mathbb{R}^m\) and that \(a \in S\).

Definition 2.4. The function \(F\) is \textbf{continuous} at \(a\) if and only if \(\lim_{x \to a} F(x) = F(a)\).

In other words, a function \(F\) is continuous at \(a\) if and only if two quantities are equal: the value of \(F\) and \(a\), and the limit of \(F(x)\) as \(x\) approaches \(a\). Sometimes one’s intuition about continuity can be misleading. For example in two variables, one might think that if the limits along horizontal and vertical lines give the value of the function, then the function is continuous. However this is not the case, as the following example shows:

Define \(F : \mathbb{R}^2 \rightarrow \mathbb{R}\) by the formulas:
\[
F(x, y) = \begin{cases} 
0 & \text{if } (x, y) = (0, 0), \\
xy[x^2 + y^2]^{-1} & \text{if } (x, y) \neq (0, 0).
\end{cases}
\]

The the value of \(F\) at the origin is zero, and the restriction of \(F\) to either the \(x\)-axis or the \(y\)-axis is also identically zero. Thus the limits as we approach \((0, 0)\) either horizontally or vertically give the value of \(F\) at \((0, 0)\). However, if we approach \((0, 0)\) along the straight line \(y = \lambda x\) with \(\lambda \neq 0\) we see that
\[
\lim_{t \to 0} F(t, \lambda t) = \lim_{t \to 0} \frac{t \lambda t}{t^2 + (\lambda t)^2} = \lim_{t \to 0} \frac{\lambda^2}{t^2 + \lambda^2 t^2} = \frac{\lambda}{1 + \lambda^2} \neq 0.
\]
Thus the limit along different straight lines give different values, and so \(F\) is not continuous at \((0, 0)\).
We have the following basic facts about limits and continuity.

**Theorem 2.5.** Suppose that $F_1, F_2 : \mathbb{R}^n \to \mathbb{R}^n$, that $a \in \mathbb{R}^n$, and that

$$\lim_{x \to a} F(x) = b_1 \quad \text{and} \quad \lim_{x \to a} F_2(x) = b_2.$$ 

Then

(a) If $\alpha_1, \alpha_2$ are scalars, then $\lim_{x \to a} (\alpha_1 F_1 + \alpha_2 F_2)(x) = \alpha_1 b_1 + \alpha_2 b_2$.

(b) $\lim_{x \to a} \langle F_1(x), F_2(x) \rangle = \langle b_1, b_2 \rangle$.

(c) If $m = 1$ so that $F_1$ and $F_2$ are real-valued and $b_1, b_2 \in \mathbb{R}$, and if $b_2 \neq 0$, then $\lim_{x \to a} \frac{F_1(x)}{F_2(x)} = \frac{b_1}{b_2}$.

**Proof of (b).** Let $\epsilon > 0$. We need to find $\delta > 0$ so that

$$||x - a|| < \delta \quad \implies \quad ||\langle F_1(x), F_2(x) \rangle - \langle b_1, b_2 \rangle|| < \epsilon.$$ 

Since we know that $\lim_{x \to a} F(x) = b_1$ and $\lim_{x \to a} F_2(x) = b_2$, we can make $||F_1(x) - b_1||$ and $||F_2(x) - b_2||$ as small as we like by making $||x - a||$ small, we work backwards and rewrite $||\langle F_1(x), F_2(x) \rangle - \langle b_1, b_2 \rangle||$ in terms of $||F_1(x) - b_1||$ and $||F_2(x) - b_2||$. In the following, we add and subtract the same quantity:

$$||\langle F_1(x), F_2(x) \rangle - \langle b_1, b_2 \rangle|| = ||\langle F_1(x) - b_1 + b_1, F_2(x) \rangle - \langle b_1, b_2 \rangle||$$

$$= ||\langle F_1(x) - b_1, F_2(x) \rangle + \langle b_1, F_2(x) \rangle - \langle b_1, b_2 \rangle||$$

$$= ||\langle F_1(x) - b_1, F_2(x) - b_2 + b_2 \rangle + \langle b_1, F_2(x) - b_2 \rangle||$$

$$= ||\langle F_1(x) - b_1, F_2(x) - b_2 \rangle + \langle F_1(x) - b_1, b_2 \rangle + \langle b_1, F_2(x) - b_2 \rangle||.$$

Note that each term on the right hand side involves either $F_1(x) - b_1$ or $F_2(x) - b_2$ or both. We now use the triangle inequality, and then the Schwarz inequality to obtain:

$$||\langle F_1(x), F_2(x) \rangle - \langle b_1, b_2 \rangle|| \leq ||F_1(x) - b_1|| ||F_2(x) - b_2|| + ||F_1(x) - b_1|| ||b_2|| + ||b_1|| ||F_2(x) - b_2||$$

If each of the three terms on the right hand side are less than $\frac{1}{3} \epsilon$, then the right hand side is less than $\epsilon$. But this is now easy to achieve. Suppose, for example that $||b_1|| \neq 0$ and $||b_2|| \neq 0$. Then simply choose $\delta > 0$ so that

$$||x - a|| < \delta \quad \implies \quad \left\{ \begin{array}{l} ||F_1(x) - b_1|| < \min \left\{ \sqrt{\frac{\epsilon}{||b_1||}}, \sqrt{\frac{\epsilon}{||b_2||}} \right\} , \\
||F_2(x) - b_2|| < \min \left\{ \sqrt{\frac{\epsilon}{||b_1||}}, \sqrt{\frac{\epsilon}{||b_2||}} \right\} . \end{array} \right.$$ 

As an exercise, you should figure out what should be done if either $||b_1|| = 0$ or $||b_2|| = 0$. Once this is done, this completes the proof of (b). You should also try to work out the proofs of (a) and (c). \qed

**Corollary 2.6.** Suppose that $F_1, F_2 : \mathbb{R}^n \to \mathbb{R}^n$, that $a \in \mathbb{R}^n$, and that $F_1$ and $F_2$ are continuous at $a$. Then

(a) If $\alpha_1, \alpha_2$ are scalars, then $(\alpha_1 F_1 + \alpha_2 F_2)$ is also continuous at $a$.

(b) $\langle F_1(x), F_2(x) \rangle$ is continuous at $a$.

(c) If $m = 1$ so that $F_1$ and $F_2$ are real-valued, and if $F_2(a) \neq 0$, then $\frac{F_1(x)}{F_2(x)}$ is continuous at $a$.

We now show that linear transformations are continuous.

**Theorem 2.7.** Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Then $T$ is continuous at every point $a \in \mathbb{R}^n$. 
Proof. Let \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \). Since \( T \) is linear, there is an \( n \times m \) matrix \( \{t_{j,k}\} \) so that \( T(x) = y = (y_1, \ldots, y_m) \) means that

\[
\begin{bmatrix}
t_{1,1} & t_{1,2} & \cdots & t_{1,n} \\
t_{2,1} & t_{2,2} & \cdots & t_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
t_{j,1} & t_{j,2} & \cdots & t_{j,n} \\
t_{m,1} & t_{m,2} & \cdots & t_{m,n}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix} =
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_m
\end{bmatrix}.
\]

Thus for \( 1 \leq j \leq n \) we have

\[ y_j = t_{j,1}x_1 + t_{j,2}x_2 + \cdots + t_{j,n}x_n = \langle T_j, x \rangle \]

where \( T_j = (t_{j,1}, t_{j,2}, \ldots, t_{j,n}) \) is the \( j^{th} \) row of the matrix. Using the Cauchy-Schwarz inequality, we then have

\[ |y_j|^2 = |\langle T_j, x \rangle|^2 \leq \|T_j\|^2 \|x\|^2 = \left( \sum_{k=1}^{n} |t_{j,k}|^2 \right) \|x\|^2. \]

If we now sum over the indices \( 1 \leq j \leq m \) we get

\[ \|y\|^2 = \sum_{j=1}^{m} |y_j|^2 \leq \left( \sum_{j=1}^{m} \sum_{k=1}^{n} |t_{j,k}|^2 \right) \|x\|^2, \]

or equivalently

\[ \|y\| = \|T(x)\| \leq C \|x\|, \]

where

\[ C = \sqrt{\left( \sum_{k=1}^{n} |t_{j,k}|^2 \right)}. \]

Note that the number \( C \) depends only on the linear transformation \( T \) and not on the point \( x \).

We now show that \( T \) is continuous at the point \( \mathbf{0} \). Let \( \epsilon > 0 \). Then if we choose \( \delta = C^{-1} \epsilon \), and if \( \|x - \mathbf{0}\| = \|x\| < \delta \), then

\[ \|T(x) - \mathbf{0}\| = \|T(x)\| \leq C \|x\| < C \delta = C C^{-1} \epsilon = \epsilon. \]

This shows that \( \lim_{x \to \mathbf{0}} T(x) = \mathbf{0} = T(0) \), so \( T \) is continuous at \( \mathbf{0} \). Finally, it is now easy to show that \( T \) is continuous at any point \( \mathbf{a} \in \mathbb{R}^n \). Let \( \epsilon > 0 \), and again choose \( \delta = C^{-1} \epsilon \) where \( C = \sqrt{\left( \sum_{k=1}^{n} |t_{j,k}|^2 \right)} \). Then since \( T \) is linear and continuous at \( \mathbf{0} \) we have

\[ \|x - \mathbf{a}\| < \delta \quad \implies \quad \|T(x - \mathbf{a})\| < \epsilon \quad \implies \quad \|T(x) - T(\mathbf{a})\| < \epsilon. \]

Thus \( \lim_{x \to \mathbf{a}} T(x) = T(\mathbf{a}) \) so \( T \) is continuous at \( \mathbf{a} \). This completes the proof. \( \Box \)

3. The derivative in several variables

We now turn to the precise definition of what it means for a function of several variables to be differentiable at a point. Let \( S \subset \mathbb{R}^n \) be an open set, so that for every \( x \in S \) there is a ball \( B(x, r) \subset S \). Let \( F : S \to \mathbb{R}^m \).

**Definition 3.1.** Let \( \mathbf{a} \in S \) and suppose \( B(\mathbf{a}, r) \subset S \). The mapping \( F \) is differentiable at the point \( \mathbf{a} \in S \) if and only if there is a linear mapping \( DF_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}^m \) so that for every \( \mathbf{h} \in \mathbb{R}^n \) with \( \|\mathbf{h}\| < r \) we have

\[ F(\mathbf{a} + \mathbf{h}) = F(\mathbf{a}) + DF_{\mathbf{a}}[\mathbf{h}] + E(\mathbf{a}, \mathbf{h}) \]

where the “error term” \( E(\mathbf{a}, \mathbf{h}) \) satisfies

\[ \lim_{\mathbf{h} \to \mathbf{0}} \frac{\|E(\mathbf{a}, \mathbf{h})\|_m}{\|\mathbf{h}\|_n} = 0. \]

That is, \( F \) is differentiable at a point \( \mathbf{a} \) if we can approximate \( F \) near \( \mathbf{a} \) by the constant mapping \( F(\mathbf{a}) \) plus a linear mapping of the change, plus an error which goes to zero faster than the size of the change.
Proposition 3.2. If a mapping $F : S \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a point $a \in S$, then $F$ is continuous at the point $a$.

Proof. Suppose that $F$ is differentiable at $a$ so that

$$F(a + h) = F(a) + DF_a[h] + E(a, h)$$

with $\lim_{h \to 0} \|E(a, h)\|_n = 0$. Then

$$\|F(a + h) - F(a)\|_m \leq \|DF_a[h]\|_m + \|E(h, a)\|_m$$

$$\leq \|DF_a\| \|h\|_n + \|E(h, a)\|_m$$

$$= \|h\|_n (\|DF_a[h]\|_m + \|E(h, a)\|_m)$$

Since the right hand side clearly goes to zero as $\|h\|_n$ goes to zero, it follows that $\lim_{x \to a} F(x) = F(a)$, and so $F$ is continuous. This completes the proof.

Proposition 3.3. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then $T$ is differentiable at every point $a \in \mathbb{R}^n$ and $DT_a = T$.

Proof. We have $T(a + h) = T(a) + T(h) = T(a) + T(h) + 0$, so $E(a, h) = 0$.

Examples:

(A) Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $F(x, y) = (x^2 y, x - 3y^2)$. Let $(a, b) \in \mathbb{R}^2$. Then

$$F(a + h, b + k) = ((a + h)^2(b + k), (a + h) - 3(b + k)^2)$$

$$= ((a^2 + 2ah + h^2)(b + k), (a + h) - 3b^2 + 2bk + k^2))$$

$$= (a^2b + a^2 k + 2ab h + 2a hk + b h^2 + h^2 k, a - 3b^2 + h - 6b k - 3k^2)$$

$$= (a^2 b, a - 3b^2) + (a^2 k + 2ab h, h - 6b k) + (2a hk + b h^2 + h^2 k, -k^2)$$

$$= F(a, b) + \begin{bmatrix} 2ab & a^2 \\ 1 & -6a \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix} + E((a, b), (h, k))$$

where $E((a, b), (h, k)) = (2a hk + b h^2 + h^2 k, -k^2)$. We let $DF_{(a,b)}$ be the linear transformation given by multiplication by the matrix $\begin{bmatrix} 2ab & a^2 \\ 1 & -6a \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix}$; and we observe that since $||(h, k)||^2 = h^2 + k^2$ we have

$$\|E((a, b), (h, k))\|^2 = (2ahk + bh^2)^2 + (-3k^2)^2$$

$$\leq (a(h^2 + k^2))^2 + b(h^2 + k^2)^2 + 9(h^2 + k^2)^2$$

$$= ||(h, k)||^2[a^2 + b^2 + 9](h^2 + k^2)$$

This shows that $F$ is differentiable at any point $(a, b)$.

(B) Consider the mapping from $\mathbb{R}^2$ to $\mathbb{R}$ given by $G(x, y) = \sin(xy)$. Then remembering that

$$|\sin(t) - t| \leq Ct^2,$$

$$|\cos(t) - 1| \leq Ct^2,$$

we have

$$G(x + h, y + k) = \sin(xy + xk + yh + hk)$$

$$= \sin(xy) \cos(xk + yh + hk) + \cos(xy) \sin(xk + yh + hk)$$

$$= \sin(xy) + \sin(xy)[\cos(xk + yh + hk) - 1] + \cos(xy) \sin(xk + yh + hk)$$

$$= \sin(xy) + \left[ y \cos(x, y) h + x \cos(x, y) k \right] + E((x, y), (h, k)).$$
3.1. Partial Derivatives.

Set \( S \subset \mathbb{R}^n \) be an open set and let \( f : S \to \mathbb{R} \). Let \( \mathbf{a} = (a_1, \ldots, a_n) \in S \). We define the partial derivatives of \( f \) at the point \( \mathbf{a} \) as follows. For \( 1 \leq j \leq n \),

\[
\frac{\partial f}{\partial x_j}(a_1, \ldots, a_n) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_{j-1}, a_j + h, a_{j+1}, \ldots, a_n) - f(a_1, \ldots, a_{j-1}, a_j, a_{j+1}, \ldots, a_n)}{h}
\]  

(3.1.1)

if this limit exists. Thus the \( j^{th} \) partial derivative of \( f \) is just the ordinary derivative of the function of one variable obtained by keeping all the variables in \( \mathbb{R}^n \) fixed except the \( j^{th} \). In computing partial derivatives, one just treats all the variables except the \( j^{th} \) as constants.

**Example 1:** Let \( f(x, y, z) = \exp[xy + yz^2] + xy \cos(yz) \). Then

\[
\frac{\partial f}{\partial x}(x, y, z) = y \exp[xy + yz^2] + y \cos(zy), \\
\frac{\partial f}{\partial y}(x, y, z) = (x + z^2) \exp[xy + yz^2] + x \cos(zy) - xyz \sin(zy), \\
\frac{\partial f}{\partial z}(x, y, z) = 2yz \exp[xy + yz^2] + xy \cos(zy) - y^2 \sin(z).
\]

**Example 2:** Let \( g(x, y) = \begin{cases} 0 & \text{if } (x, y) = (0, 0), \\
(x^2 + y^2) \sin((x^2 + y^2)^{-1}) & \text{if } (x, y) \neq (0, 0). \end{cases} \)

Then when \( (x, y) \neq (0, 0) \) we have

\[
\frac{\partial g}{\partial x}(x, y) = 2y \sin((x^2 + y^2)^{-1}) - \frac{2x}{(x^2 + y^2)^2} \cos((x^2 + y^2)^{-1}), \\
\frac{\partial g}{\partial y}(x, y) = 2x \sin((x^2 + y^2)^{-1}) - \frac{2y}{(x^2 + y^2)^2} \cos((x^2 + y^2)^{-1}).
\]

On the other hand, when \( (x, y) = (0, 0) \),

\[
\frac{\partial g}{\partial x}(0, 0) = \lim_{h \to 0} \frac{g(h, 0) - g(0, 0)}{h} = \lim_{h \to 0} \frac{h^2 \sin(h^{-2})}{h^2} = 0, \\
\frac{\partial g}{\partial y}(0, 0) = \lim_{k \to 0} \frac{g(0, k) - g(0, 0)}{k} = \lim_{k \to 0} \frac{k^2 \sin(k^{-2})}{k} = 0.
\]

Note that if \( f : S \to \mathbb{R} \) and if the partial derivative with respect to \( x_j \) of \( f \) exist at every point of \( S \), then \( \frac{\partial f}{\partial x_j} : S \to \mathbb{R} \) is another function, and we can try to take its partial derivatives. Thus one can define higher order partial derivatives. We will see later that for most reasonable functions, it does not matter in which order one takes the derivatives. Thus if \( m_1 + \cdots + m_n = M \), one can define partial derivatives of order \( M \) which we write

\[
\frac{\partial^{m_1+\cdots+m_n} f}{\partial x_1^{m_1} \cdots \partial x_n^{m_n}}(x_1, \ldots, x_n).
\]

This means we take \( m_1 \) derivatives in \( x_1 \), \( m_2 \) derivatives in \( x_2 \), and so on.

**Example 3:** Let \( h(x, y) = \exp[x^2y^3] \). Then

\[
\frac{\partial h}{\partial x}(x, y) = 2x \exp[x^2y^3], \\
\frac{\partial^2 h}{\partial x^2}(x, y) = 2 \exp[x^2y^3] + 4x^2 \exp[x^2y^3], \\
\frac{\partial^2 h}{\partial y \partial x}(x, y) = 6xy^2 \exp[x^2y^3], \\
\frac{\partial h}{\partial y}(x, y) = 3y^2 \exp[x^2y^3], \\
\frac{\partial^2 h}{\partial x \partial y}(x, y) = 6xy^2 \exp[x^2y^3], \\
\frac{\partial^2 h}{\partial y^2}(x, y) = 6y \exp[x^2y^3] + 9y^4 \exp[x^2y^3].
\]
3.2. The Jacobian matrix.

We now turn to the problem of calculating the linear mapping $DF_a$ for a differentiable function $F : \mathbb{R}^n \to \mathbb{R}^m$. Recall that if $F$ is differentiable at $a$, then for all vectors $h \in \mathbb{R}^n$ with sufficiently small length we have

$$F(a + h) = F(a) + DF_a[h] + E(a, h)$$

(3.2.1)

where

$$\lim_{h \to 0} \frac{\|E(a, h)\|_m}{\|h\|_n} = 0.$$ 

We should be able to recognize the linear transformation $DF_a$ as an $m \times n$ matrix.

**Theorem 3.4.** If $F : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a$, if $F(x) = (f_1(x), \ldots, f_m(x))$, and if $h = (h_1, \ldots, h_n)$, then

$$DF_a[h] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}.$$ 

**Proof.** If $F$ is differentiable at $a$, then there is a linear transformation $DF_a : \mathbb{R}^n \to \mathbb{R}^m$ so that equation (3.2.1) holds. This transformation is then given by and $m \times n$ matrix, say

$$\begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix}.$$ 

Our job is to figure out what the entries $A_{j,k}$ are.

Let us consider what happens in equation (3.2.1) if we take $h = (h,0,\ldots,0)$, so that $h$ is just a scalar multiple $h \neq 0$ of the first basis element $e_1$. Since

$$F(x) = (f_1(x_1, x_2, \ldots, x_n), f_2(x_1, x_2, \ldots, x_n), \ldots, f_m(x_1, x_2, \ldots, x_n)),$$

equation (3.2.1) says

$$\left( f_1(a_1 + h, \ldots, a_n), f_2(a_1 + h, \ldots, a_n), \ldots, f_m(a_1 + h, \ldots, a_n) \right)$$

$$= \left( f_1(a_1, \ldots, a_n), f_2(a_1, \ldots, a_n), \ldots, f_m(a_1, \ldots, a_n) \right)$$

$$+ \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix} \begin{bmatrix} h \\ 0 \\ \vdots \\ 0 \end{bmatrix} + E(a, he_1)$$

$$= \left( f_1(a_1, \ldots, a_n), f_2(a_1, \ldots, a_n), \ldots, f_m(a_1, \ldots, a_n) \right)$$

$$+ \left( A_{1,1} h, A_{2,1} h, \ldots, A_{m,1} h \right) + E(a, he_1)$$

$$= \left( f_1(a_1, \ldots, a_n), f_2(a_1, \ldots, a_n), \ldots, f_m(a_1, \ldots, a_n) \right)$$

$$+ h \left( A_{1,1}, A_{2,1}, \ldots, A_{m,1} \right) + E(a, he_1).$$

Rearranging this equation and then dividing by $h$ then gives

$$\left( \frac{f_1(a_1 + h, \ldots, a_n) - f_1(a_1, \ldots, a_n)}{h}, \ldots, \frac{f_m(a_1 + h, \ldots, a_n) - f_m(a_1, \ldots, a_n)}{h} \right)$$

$$= \left( A_{1,1}, A_{2,1}, \ldots, A_{m,1} \right) + \frac{E(a, he_1)}{h}.$$ 

Taking the limit as $h \to 0$ now shows that

$$A_{j,k} = \lim_{h \to 0} \frac{f_j(a_1 + h, \ldots, a_n) - f_j(a_1, \ldots, a_n)}{h} = \frac{\partial f_j}{\partial x_1}(a).$$
Since we can do a similar argument for any other coefficients, this completes the proof. \hfill \Box

11/16/12

3.3. Real-valued functions and the gradient.

Suppose \( S \subset \mathbb{R}^n \) is open and that \( f : S \subset \mathbb{R}^n \to \mathbb{R} \) is differentiable at a point \( a \in S \). Then for all vectors \( h = (h_1, \ldots, h_n) \) with sufficiently small length, we have

\[
\begin{align*}
f(a + h) &= f(a) + \left[ \frac{\partial f}{\partial x_1}(a) \ h_1 \quad \frac{\partial f}{\partial x_2}(a) \ h_2 \quad \cdots \quad \frac{\partial f}{\partial x_n}(a) \ h_n \right] + E(a, h) \\
&= f(a) + \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) h_j + E(a, h)
\end{align*}
\]

where \( \lim_{h \to 0} \frac{|E(a, h)|}{||h||} = 0. \)

**Definition 3.5.** If \( f \) is a real-valued function differentiable at a point \( a \in \mathbb{R}^n \) the gradient of \( f \) at \( a \) is the vector

\[
\nabla f(a) = \left( \frac{\partial f}{\partial x_1}(a), \frac{\partial f}{\partial x_2}(a), \ldots, \frac{\partial f}{\partial x_n}(a) \right).
\]

We can then rewrite (3.3.1) as

\[
f(a + h) = f(a) + \nabla f(a) \cdot h + E(a, h).
\]

We now use this to find the geometric significance of the grandaunt \( \nabla f(a) \). Consider the restriction of \( f \) to a straight line passing through the point \( a \). We can parameterize this line as a function of a real variable \( t \) by considering the vector \( a + te \) where \( e \) is some unit vector. Then the restriction of \( f \) to this line, as a function of \( t \) is given by \( g(t) = f(a + te) \). Note that the ordinary one-variable derivative \( g'(0) \) is then the rate of change of the function \( f \) in the direction of the vector \( e \). Let us compute this derivative:

\[
g'(0) = \lim_{t \to 0} \frac{g(t) - g(0)}{t} = \lim_{t \to 0} \frac{1}{t} [f(a + te) - f(a)] = \lim_{t \to 0} \frac{1}{t} [\nabla f(a) \cdot te + E(a, te)]
\]

\[
= \nabla f(a) \cdot e + \lim_{t \to 0} \frac{E(a, te)}{t} = \nabla f(a) \cdot e.
\]

We thus have the following:

**Proposition 3.6.** The gradient of \( f \) at \( a \) points in the direction in which \( f \) increases most rapidly, and the length of the gradient is this fastest rate of increase.

Again, let \( f : S \to \mathbb{R} \) where \( S \subset \mathbb{R}^n \) is an open subset. If \( a \in S \), the level set of \( f \) passing through \( a \) is the subset of \( S \) given by

\[
S = \{ x \in S : f(x) = f(a) \}.
\]

For example, if \( f : \mathbb{R}^3 \to \mathbb{R} \) is given by \( f(x, y, z) = x^2 + y^2 + z^2 \), then for any \( (a, b, c) \in \mathbb{R}^3 \), the level set of \( f \) passing through this point,

\[
S = \{ (x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = a^2 + b^2 + c^2 \},
\]

is the sphere centered at the origin with radius \( \sqrt{a^2 + b^2 + c^2} \). Note that if \( (a, b, c) = (0, 0, 0) \), this level set is just a single point. However, most of the time, the level set is a 2-dimensional surface passing through the point. In general, if \( f : S \subset \mathbb{R}^n \to \mathbb{R} \), if \( a \in S \), and if \( \nabla f(a) \neq 0 \), then the level set of \( f \) passing through the point \( a \) will be an \( (n-1) \)-dimensional surface. (For a proof of this, take Math 621.)

Suppose that \( \nabla f(a) \neq 0 \) and that \( S \) is the level set of \( f \) passing through \( a \). Suppose \( \Phi = (\varphi_1, \ldots, \varphi_n) : (-1, +1) \to \mathbb{R}^n \) is a smooth curve such that \( \Phi(0) = a \), such that \( \Phi(t) \) is a point of \( S \) for \(-1 < t < 1 \), and such that \( \Phi(t) \neq \Phi(a) \) if \( t \neq 0 \). This means that
(a) \( \Phi(0) = a \), and
(b) \( f(\Phi(t)) = 0 \) for \(-1 < t < +1\),
(c) \( \Phi(t) \neq a \) for \(0 < |t| < 1\).

Let \(0 < |t| < 1\) and let us again use equation (3.3.2) with \( h = \Phi(t) - \Phi(0) \). We have
\[
\begin{align*}
f(\Phi(t)) &= f(\Phi(0) + [\Phi(t) - \Phi(0)]) \\
&= f(\Phi(0)) + \nabla f(\Phi(0)) \cdot [\Phi(t) - \Phi(0)] + E(\Phi(t) - \Phi(0)).
\end{align*}
\]
But since \( \Phi(t) \in S \), \( f(\Phi(t)) = f(a) \), and we can cancel this from both sides of the equation. Thus we get
\[
\nabla f(a) \cdot [\Phi(t) - \Phi(0)] = -E(\Phi(t) - \Phi(0)).
\]
For \( t \neq 0 \), we can divide both sides of this equation by \( t \):
\[
\nabla f(a) \cdot \left[ \frac{\Phi(t) - \Phi(0)}{t} \right] = -\frac{1}{t}E(\Phi(t) - \Phi(0)).
\]
Now let \( t \to 0 \). For the left hand side we have
\[
\lim_{t \to 0} \left[ \frac{\Phi(t) - \Phi(0)}{t} \right] = \lim_{t \to 0} \left[ \frac{\varphi_1(t) - \varphi_1(0)}{t}, \ldots, \frac{\varphi_n(t) - \varphi_n(0)}{t} \right] = \left[ \lim_{t \to 0} \frac{\varphi_1(t) - \varphi_1(0)}{t}, \ldots, \lim_{t \to 0} \frac{\varphi_n(t) - \varphi_n(0)}{t} \right] = [\varphi'_1(0), \ldots, \varphi'_n(0)].
\]
This is a vector \( \Phi'(0) \) which is tangent to the level surface \( S \). On the right hand side, if we take the length we have
\[
\left| \frac{1}{t}E(\Phi(t) - \Phi(0)) \right| = \frac{1}{|t|} \left| E(\Phi(t) - \Phi(0)) \right| = \frac{\left| E(\Phi(t) - \Phi(0)) \right| \left| \Phi(t) - \Phi(0) \right|}{\left| \Phi(t) - \Phi(0) \right|} = \left| \Phi(t) - \Phi(0) \right|.
\]
Letting \( t \to 0 \), the first term in this product goes to zero by the properties of the error term in the definition of differentiability, and the second term goes to the length of the vector \( \Phi'(0) \). It follows that
\[
\nabla f(a) \cdot \Phi'(0) = 0
\]
and so the gradient \( \nabla f(a) \) must be perpendicular to the vector \( \Phi'(0) \). However, the vector \( \Phi'(0) \) is really an arbitrary vector which is tangent to the level surface of \( f \) at \( a \). Thus

**Proposition 3.7.** If \( \nabla f(a) \neq 0 \), the vector \( \nabla f(a) \) is perpendicular to the level surface
\[
S = \{ x \in S : f(x) = f(a) \}
\]
**passing through** \( a \).

We can use this result to find the formula of the tangent plane to the level surface \( S \). Thus \( x \) belongs to this plane if and only if the vector \( x - a \) is perpendicular to the gradient \( \nabla f(a) \). Thus the equation of the tangent plane is
\[
\nabla f(a) \cdot [x - a] = 0, \quad \text{or equivalently} \quad \nabla f(x) \cdot x = \nabla f(a) \cdot a.
\]

**Example:** Find the equation of the tangent plane to the graph of the function \( z = f(x,y) \) at the point \((a, b, f(a,b))\).

The graph of \( f \) is
\[
G_f = \{(x,y,z) \in \mathbb{R}^3 : z = f(x,y)\} = \{(x,y,z) \in \mathbb{R}^3 : f(x,y) - z = 0\}.
\]
Define $F : \mathbb{R}^3 \to \mathbb{R}$ by $F(x, y, z) = f(x, y) - z$. Then the graph of $f$ is just the level surface of the function $f(x, y, z) = 0$. The point $(a, b, f(a, b))$ lies on this surface. Since
\[
\nabla F(x, y, z) = \left( \frac{\partial F}{\partial x}(x, y, z), \frac{\partial F}{\partial y}(x, y, z), \frac{\partial F}{\partial z}(x, y, z) \right)
\]
we have
\[
\nabla F(x, y, z) = \left( \frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y), -1 \right),
\]
the equation of the tangent plane to the graph of $f$ at $(a, b, f(a, b))$ is
\[
\frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) + (-1)(z - f(a, b)),
\]
or equivalently
\[
\begin{align*}
z &= f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b).
\end{align*}
\]

3.4. The Chain Rule.

**Theorem 3.8.** Suppose that $a \in \mathbb{R}^n$, that $F : U \subset \mathbb{R}^n \to V \subset \mathbb{R}^m$ is defined in an open set $U$ containing $a$ with values in an open set $V$, and that $F$ is differentiable at $a$. Suppose also that $G : V \subset \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $F(a)$. Then $G \circ F : S \subset \mathbb{R}^n \to \mathbb{R}^p$ is also differentiable at $a$. Moreover
\[
D(G \circ F)_a = DG_{F(a)} \circ DF_a.
\]

**Proof.** For $h \in \mathbb{R}^n$ sufficiently small we have
\[
(G \circ F)(a + h) = G(F(a) + h)
\]
\[
= G(F(a)) + DG_{F(a)}[F(a + h) - F(a)] + \|E_F(a, h)\|_p
\]
\[
= G(F(a)) + DG_{F(a)}[DF_a(h) + EF(a, h)] + \|E_F(a, h)\|_p
\]
\[
= G(F(a)) + DG_{F(a)}[DF_a(h)] + DG_{F(a)}[EF(a, h)] + \|E_F(a, h)\|_p
\]
\[
= G \circ F(a) + [DG_{F(a)} \circ DF_a] h + \tilde{E}(a, h)
\]
where
\[
\tilde{E}(a, h) = DF_a[EF(a, h)] + \|E_F(a, h)\|_p
\]
\[
= DF_a[EF(a, h)] + \|E_F(a, h)\|_p + EF(a, h) + EF(a, h)
\]
If we can show that
\[
\lim_{h \to 0} \frac{\|\tilde{E}(a, h)\|_p}{\|h\|_n} = 0
\]
the we will have proved the theorem. Using the triangle inequality, it suffices to show that
\[
\lim_{h \to 0} \frac{\|DF_a[EF(a, h)]\|_p}{\|h\|_n} = 0 \quad \text{and} \quad \lim_{h \to 0} \frac{\|E_G(F(a), DF_a[h] + EF(a, h))\|_p}{\|h\|_n} = 0.
\]
The first limit is not hard since we have
\[
\|DF_a[EF(a, h)]\|_p \leq C \|EF(a, h)\|_m
\]
and we can then use the fact that $\lim_{h \to 0} \|EF(a, h)\|_m \|h\|_n^{-1} = 0$. To show that the second limit exists and equals zero, let $0 < \epsilon < 1$, and then let $0 < \delta < \epsilon$ be such that
\[
\|v\|_m < \delta \implies \|E_G(F(a), v)\|_p < \epsilon \|v\|_m,
\]
\[
\|w\|_n < \delta \implies \|E_F(a, w)\|_m < \frac{1}{2} \epsilon \|w\|_n.
\]

Now
\[
\|DF_a[h] + E_F(a, h)\|_m \leq \|DF_a[h]\|_m - \|E_F(a, h)\|_m
\]
\[
\leq \||DF_a|| \|h\|_n + \|E_F(a, h)\|_m
\]
Thus if \( ||h||_n < \min \left\{ 2 ||DF_a||^{-1} \delta, 2^{-1} \delta \right\} \) we have

\[
||DF_a[h] + EF(a, h)||_m \leq ||DF_a|| \cdot ||h||_n + ||EF(a, h)||_m < \frac{1}{2} \delta + \frac{1}{2} \epsilon ||h|| < \delta.
\]

Then it follows that

\[
||EG(F(a), DF_a[h] + EF(a, h))||_p < \epsilon (||DF_a[h] + EF(a, h)||_m < \epsilon \cdot ||DF_a|| \cdot ||h||_n + ||EF(a, h)||_m < \epsilon \cdot ||DF_a|| \cdot ||h||_n + \frac{1}{2} \epsilon \cdot ||h||_n).
\]

and this gives the desired result. \( \square \)

3.5. Formulas for the Chain Rule. 11/26/12

Let \( G : \mathbb{R}^m \to \mathbb{R}^n \) and \( F : \mathbb{R}^n \to \mathbb{R}^p \), with \( G \) differentiable at \( a \in \mathbb{R}^n \) and \( F \) differentiable at \( b = F(a) \in \mathbb{R}^m \). Let us denote the variables in \( \mathbb{R}^m \) by \( t = (t_1, \ldots, t_m) \), the variables in \( \mathbb{R}^n \) by \( x = (x_1, \ldots, x_n) \), and the variables in \( \mathbb{R}^p \) by \( y = (y_1, \ldots, y_p) \). Let

\[
F(x) = F(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_p(x_1, \ldots, x_n)),
\]

\[
G(t) = G(t_1, \ldots, t_m) = (g_1(t_1, \ldots, t_m), \ldots, g_n(t_1, \ldots, t_m)).
\]

We can think of this as \( y = F(x) \) and \( x = G(t) \). We denote the composition \( F \circ G \) by \( H : \mathbb{R}^m \to \mathbb{R}^p \). Thus

\[
H(t) = H(t_1, \ldots, t_m)
\]

\[
= F(G(t_1, \ldots, t_m))
\]

\[
= (f_1(G(t_1, \ldots, t_m)), \ldots, f_p(G(t_1, \ldots, t_m)))
\]

\[
= \left( f_1(g_1(t_1, \ldots, t_m), \ldots, g_n(t_1, \ldots, t_m)), \ldots, f_p(g_1(t_1, \ldots, t_m), \ldots, g_n(t_1, \ldots, t_m)) \right)
\]

\[
= (h_1(t_1, \ldots, t_m), \ldots, h_p(t_1, \ldots, t_m)).
\]

We want to compute the partial derivatives of \( H \), such as \( \frac{\partial h_j}{\partial t_k}(t_1, \ldots, t_m) \). According to the Chain Rule, the linear mapping \( DH_a \) is equal to the composition \( DF_{G(a)} \circ DG_a \). Moreover, we have seen that each of these mappings is given by multiplication by the appropriate Jacobian matrix. We have

\[
DF_{G(a)} = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1}(G(a)) & \frac{\partial f_1}{\partial x_2}(G(a)) & \cdots & \frac{\partial f_1}{\partial x_n}(G(a)) \\
\frac{\partial f_2}{\partial x_1}(G(a)) & \frac{\partial f_2}{\partial x_2}(G(a)) & \cdots & \frac{\partial f_2}{\partial x_n}(G(a)) \\
& \ddots & \ddots & \ddots \\
\frac{\partial f_p}{\partial x_1}(G(a)) & \frac{\partial f_p}{\partial x_2}(G(a)) & \cdots & \frac{\partial f_p}{\partial x_n}(G(a))
\end{bmatrix}
\]

(a \( p \times n \) matrix)

\[
DG_a = \begin{bmatrix}
\frac{\partial g_1}{\partial t_1}(a) & \frac{\partial g_1}{\partial t_2}(a) & \cdots & \frac{\partial g_1}{\partial t_m}(a) \\
\frac{\partial g_2}{\partial t_1}(a) & \frac{\partial g_2}{\partial t_2}(a) & \cdots & \frac{\partial g_2}{\partial t_m}(a) \\
& \ddots & \ddots & \ddots \\
\frac{\partial g_n}{\partial t_1}(a) & \frac{\partial g_n}{\partial t_2}(a) & \cdots & \frac{\partial g_n}{\partial t_m}(a)
\end{bmatrix}
\]

(an \( n \times m \) matrix)
It follows from the one-variable Chain Rule where \( 0 \). Let

\[
\text{proof.}
\]

\[\text{We will discuss the following two results.}\]

\[\text{Let } S \subset \mathbb{R}^n \text{ be an open set and let } F : \mathbb{R}^n \rightarrow \mathbb{R}^p \text{ be a mapping. Suppose that all the first partial derivatives of the components of } F \text{ exist at every point of } S. \text{ If these partial derivatives are continuous at a point } \mathbf{a} \in S, \text{ then the mapping } F \text{ is differentiable at the point } \mathbf{a}.\]

\[\text{Proof.} \quad \text{We can write } F(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_p(\mathbf{x})) \text{ where each } f_j : S \rightarrow \mathbb{R}, \text{ and the partial derivatives } \frac{\partial f_j}{\partial x_k}(\mathbf{x}) \text{ exist at each point } \mathbf{x} \in S \text{ for } 1 \leq j \leq p, 1 \leq k \leq n. \text{ For } 1 \leq k \leq n \text{ we shall write}
\]

\[\frac{\partial F}{\partial x_k} = \left( \frac{\partial f_1}{\partial x_k}(\mathbf{x}), \ldots, \frac{\partial f_p}{\partial x_k}(\mathbf{x}) \right).\]

It follows from the one-variable Chain Rule that if \( \{ e_1, \ldots, e_n \} \) is the standard basis for \( \mathbb{R}^n \) then

\[F(\mathbf{x} + \theta \mathbf{e}_k) = F(\mathbf{x}) + \theta \left( \frac{\partial f_1}{\partial x_k}(\mathbf{x} + \theta \mathbf{e}_k), \ldots, \frac{\partial f_p}{\partial x_k}(\mathbf{x} + \theta \mathbf{e}_k) \right)\]

where \( 0 < \theta < 1 \) for \( 1 \leq l \leq n.\)

Let \( \mathbf{a} = (a_1, \ldots, a_n) \) and \( \mathbf{h} = (h_1, \ldots, h_n). \) Then we can write

\[F(\mathbf{a} + \mathbf{h}) - F(\mathbf{a}) = \sum_{j=1}^{n} F(\mathbf{a} + \sum_{l=1}^{j} h_l e_l) - F(\mathbf{a} + \sum_{l=1}^{j-1} h_l e_l)\]

\[= \sum_{j=1}^{n} \left( \frac{\partial f_1}{\partial x_k}(\mathbf{a} + \sum_{l=1}^{j} h_l e_l + \theta_{1,j} h_j e_k), \ldots, \frac{\partial f_p}{\partial x_k}(\mathbf{a} + \sum_{l=1}^{j} h_l e_l + \theta_{n,j} h_j e_k) \right)\]

\[= \sum_{j=1}^{n} h_j \left( \frac{\partial f_1}{\partial x_k}(\mathbf{a}), \ldots, \frac{\partial f_p}{\partial x_k}(\mathbf{a}) \right) + \sum_{j=1}^{n} h_j E_j(\mathbf{a}, \mathbf{h})\]

where

\[E_j(\mathbf{a}, \mathbf{h}) = \left( \frac{\partial f_1}{\partial x_k}(\mathbf{a} + \sum_{l=1}^{j-1} h_l e_l + \theta_{1,j} h_j e_k) - \frac{\partial f_1}{\partial x_k}(\mathbf{a}), \ldots, \frac{\partial f_p}{\partial x_k}(\mathbf{a} + \sum_{l=1}^{j-1} h_l e_l + \theta_{n,j} h_j e_k) - \frac{\partial f_p}{\partial x_k}(\mathbf{a}) \right)\].
Since the partial derivatives are all continuous, we have \( \lim_{h \to 0} ||E_j(a, h)||_p = 0 \). Then using the triangle inequality and the Cauchy-Schwarz inequality, we have
\[
\left\| \sum_{j=1}^n h_j E_j(a, h) \right\|_p \leq \sum_{j=1}^n |h_j| \left\| E_j(a, h) \right\|_p \leq \left( \sum_{j=1}^n |h_j|^2 \right)^{\frac{p}{2}} \left( \sum_{j=1}^n \left\| E_j(a, h) \right\|_p^2 \right)^{\frac{p}{2}}
= |h|_n \left( \sum_{j=1}^n \left\| E_j(a, h) \right\|_p^2 \right)^{\frac{p}{2}}.
\]
This shows that the error has the right estimate, and completes the proof. \( \square \)

**Theorem 3.10.** Let \( S \subset \mathbb{R}^n \) be an open set and let \( f : S \to \mathbb{R} \) be a real-valued function. Suppose that the partial derivatives \( \frac{\partial f}{\partial x_j}, \frac{\partial f}{\partial x_k}, \frac{\partial^2 f}{\partial x_j \partial x_k}, \) and \( \frac{\partial^2 f}{\partial x_k \partial x_j} \) exist at each point of \( S \). Suppose that both \( \frac{\partial^2 f}{\partial x_j \partial x_k} \) and \( \frac{\partial^2 f}{\partial x_k \partial x_j} \) are continuous at a point \( a \in S \). Then
\[
\frac{\partial^2 f}{\partial x_j \partial x_k}(a) = \frac{\partial^2 f}{\partial x_k \partial x_j}(a).
\]
(For a proof, see for example Section 8.23 in the text.)

Note the following application of Theorem 3.10. If \( G(x) = (g_1(x), \ldots, g_n(x)) \) is a function defined on \( \mathbb{R}^n \), we can ask if there is a function \( f : \mathbb{R}^n \to \mathbb{R} \) so that \( \nabla f(x) = G(x) \). That is, given \( n \) real-valued functions \( g_1, \ldots, g_n \) on \( \mathbb{R}^n \) we ask if there is a real-valued function \( f \) defined on \( \mathbb{R}^n \) so that \( \frac{\partial f}{\partial x_j}(x) = g_j(x) \). If such a function \( f \) exists, we would then have
\[
\frac{\partial g_j}{\partial x_k}(x) = \frac{\partial}{\partial x_k} \frac{\partial f}{\partial x_j}(x) = \frac{\partial^2 f}{\partial x_k \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_k}(x) = \frac{\partial g_k}{\partial x_j}(x).
\]
Thus a necessary condition for the existence of the function \( f \) is that for all \( 1 \leq j, k \leq n \) we have
\[
\frac{\partial g_j}{\partial x_k}(x) = \frac{\partial g_k}{\partial x_j}(x).
\]

3.7. **List of other topics.**

(a) Implicit differentiation;
(b) Maxima, minima, and saddle points;
(c) Second order Taylor series;
(d) Second derivative test;
(e) Constrained extrema and Lagrange multipliers.

3.8. **Finding stationary points and local extrema.**

Let \( S \subset \mathbb{R}^n \) be an open set, and let \( f : S \to \mathbb{R} \). In analogy with the concepts for functions of one variable, we make the following definitions.

**Definition 3.11.**

(a) A point \( x_0 \in S \) is a **local maximum** for \( f \) if there is an open ball \( B(x_0, r) \subset S \) so that \( f(y) \leq f(x_0) \) for all \( y \in B(x_0, r) \).

(b) A point \( x_0 \in S \) is a **local minimum** for \( f \) if there is an open ball \( B(x_0, r) \subset S \) so that \( f(y) \geq f(x_0) \) for all \( y \in B(x_0, r) \).
Theorem 3.12. Let $f : S \subset \mathbb{R}^n \to \mathbb{R}$, and suppose that the first partial derivatives of $f$ exist at a point $x_0 \in S$. If $a$ is either a local maximum or local minimum for $f$, then $\nabla f(x_0) = 0$.

Proof. Let $\{e_1, \ldots, e_n\}$ be the standard basis for $\mathbb{R}^n$. For $1 \leq j \leq n$, let $g_j(t) = f(x_0 + te_j)$. Then

$$g_j'(0) = \lim_{h \to 0} \frac{g_j(h) - g_j(0)}{h} = \lim_{h \to 0} \frac{f(x_0 + he_j) - f(x_0)}{h} = \frac{\partial f}{\partial x_j}(x_0).$$

On the other hand, if $f$ has a local maximum or local minimum at the point $x_0$, then each function $g_j$ has a local maximum or local minimum at the point $t = 0$. From 1-variable calculus, it follows that $g_j'(0) = 0$, and hence $\nabla f(x_0) = 0$. This completes the proof. 

Definition 3.13. If $f : S \to \mathbb{R}$ is differentiable, a point $x_0 \in S$ where $\nabla f(x_0) = 0$ is called a stationary point or critical point of $f$. A critical point for $f$ is called a saddle point for $f$ if for every $r > 0$ there are points $x_1, x_2 \in (x_0, r) \cap S$ such that $f(x_1) < f(x_0) < f(x_2)$.

Examples:

1) What is the difference between the critical points of the functions $f(x, y) = x^2 + y^2$, $g(x, y) = x^2 - y^2$ and $h(x, y) = x^2 - y^2$?

2) Find the critical points of the function $f(x, y) = x - 2y + \log \sqrt{x^2 + y^2} + 3 \arctan \left( \frac{y}{x} \right)$.

We compute $\nabla f(x, y)$:

$$\frac{\partial f}{\partial x}(x, y) = 1 + \frac{1}{2} \cdot \frac{2x}{x^2 + y^2} + 3 \cdot \frac{1}{1 + \left( \frac{y}{x} \right)^2} \cdot \frac{-y}{x^2} = \frac{x^2 + y^2 + x - 3y}{x^2 + y^2}$$

$$\frac{\partial f}{\partial y}(x, y) = -2 + \frac{1}{2} \cdot \frac{2y}{x^2 + y^2} + 3 \cdot \frac{1}{1 + \left( \frac{y}{x} \right)^2} \cdot \frac{1}{x} = \frac{-2x^2 - 2y^2 + y + 3x}{x^2 + y^2}$$

Thus we need to solve the (non-linear) system of two equations in two unknowns

$$x^2 + y^2 + x - 3y = 0,$$

$$-2x^2 - 2y^2 + y + 3x = 0.$$

Completing the squares, these are the equations of two circles:

$$\left( x + \frac{1}{2} \right)^2 + \left( y - \frac{3}{2} \right)^2 = \frac{5}{2},$$

$$\left( x - \frac{3}{4} \right)^2 + \left( y - \frac{1}{4} \right)^2 = \frac{5}{8}.$$  

These circles intersect at $(0, 0)$ and at $(1, 1)$. Since the function is not defined at $(0, 0)$, the only critical point is at $(1, 1)$.

3) Given data points $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, find the equation of the straight line $y = f(x) = Ax + B$ which has the best “least squares” fit to this data. That is, find $A$ and $B$ so that $E(A, B) = \sum_{j=1}^n |f(x_j) - y_j|^2$ is as small as possible.

We have

$$E(A, B) = \sum_{j=1}^n |Ax_j + B - y_j|^2 = \sum_{j=1}^n (Ax_j + B - y_j)^2.$$
Then if $\nabla E(A, B) = 0$ we have
\[
\frac{\partial E}{\partial A} = \sum_{j=1}^{n} 2(Ax_j + B - y_j)x_j = 2A \sum_{j=1}^{n} x_j^2 + 2B \sum_{j=1}^{n} x_j - 2 \sum_{j=1}^{n} x_j y_j = 0,
\]
\[
\frac{\partial E}{\partial B} + \sum_{j=1}^{n} 2(Ax_j + B - y_j) = 2A \sum_{j=1}^{n} x_j + 2nB - 2 \sum_{j=1}^{n} y_j = 0.
\]
We can now solve for $A$ and $B$.

### 3.9. Constrained Extrema and Lagrange Multipliers.

In one-variable calculus, we learn that to find the maximum or minimum of a function $f$ on a closed interval $[a, b]$, the only points where this can occur are of one of the following types:

(a) Points in the open interval $(a, b)$ where the derivative does not exist.
(b) Points in the open interval $(a, b)$ where the derivative exists and equals zero.
(c) The endpoints $a$ and $b$ of the interval.

Moreover, if there is a point $c \in (a, b)$ at which $f'(c) = 0$, we have the following test to determine if the point is a local maximum or a local minimum:

(i) If $f''(c) > 0$ then the point $c$ is a local minimum;
(ii) If $f''(c) < 0$ then the point $c$ is a local maximum;
(iii) If $f''(c) = 0$, the test is inconclusive.

So far we have developed the analogue of conditions (a) and (b) for functions of several variables. Thus if $S \subset \mathbb{R}^n$ is open, if $f : S \to \mathbb{R}$, and if $a \in S$ is either a local maximum or a local minimum for $f$, then either some first order partial derivative of $f$ does not exist at $a$ or we must have $\nabla f(a) = 0$. Thus to find local extrema of $f$ in an open set, we look for points where some partial derivative does not exist, and we look for points $x = (x_1, \ldots, x_n)$ which satisfy the system of $n$ equations in $n$ unknowns:
\[
\frac{\partial f}{\partial x_1}(x_1, \ldots, x_n) = 0, \\
\frac{\partial f}{\partial x_2}(x_1, \ldots, x_n) = 0, \\
\vdots \\
\frac{\partial f}{\partial x_n}(x_1, \ldots, x_n) = 0.
\]

To bring the theory of maxima and minima in several variables to the same point as the theory in one variable, we need to do two more things: we need to investigate what happens on the boundary of an open set, and we need to develop an analogue of the second derivative test. We begin with the first issue.

Let us begin with an example: find the maximum and minimum values of the function $f(x, y) = x^2y$ on the closed set
\[
D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.
\]
Note that the function $f$ is differentiable at every point, and
\[
\nabla f(x, y) = (2xy, x^2) = (0, 0) \iff (x, y) = (0, 0).
\]
The point $(0, 0)$ is a critical point in the interior of the set. We have $f(0, 0) = 0$, but it is clear that this point is a saddle point since $f(x, y) > 0$ if $x \neq 0$ and $y > 0$ while $f(x, y) < 0$ if $x \neq 0$ and $y < 0$. Thus the maximum and minimum can only occur on the boundary of the set $S$ which is the unit circle
\[
T = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}.
\]
When dealing with functions of a single variable, the boundary of an interval consists of only two points, but now we see that the boundary of a set in higher dimensions can consist of an infinite number of points. How do we deal with this problem?

**Method 1: Parameterize the boundary.**

We can parameterize the points \((x, y)\) on the unit circle \(T\) by \(x = \cos(\theta)\) and \(y = \sin(\theta)\) for \(0 \leq \theta \leq 2\pi\). Then on the circle, the function \(f\) becomes a function of one variable

\[
g(\theta) = f(\cos(\theta), \sin(\theta)) = \cos^2(\theta) \sin(\theta).
\]

We can then find the maximum and minimum by using one-variable techniques. We have

\[
g'(\theta) = -2 \cos(\theta) \sin^2(\theta) + \cos^3(\theta)
\]

\[
= \cos(\theta) \left[ \cos^2(\theta) - 2 \sin^2(\theta) \right]
\]

\[
= \cos(\theta) [1 - 3 \sin^2(\theta)].
\]

Thus \(g'(\theta) = 0\) when \(\cos(\theta) = 0\) or \(\sin(\theta) = \pm \sqrt{\frac{1}{3}}\). Solving, we get six possible points on \(T\) where \(f\) can be a local extremum:

\[
(0, 1), \quad (0, -1), \quad \left( \frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3} \right), \quad \left( \frac{\sqrt{3}}{3}, -\frac{\sqrt{3}}{3} \right), \quad \left( -\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3} \right), \quad \left( -\frac{\sqrt{3}}{3}, -\frac{\sqrt{3}}{3} \right).
\]

While this method sometimes works quite well, one can run into difficulties. Thus if we want to maximize or minimize \(f(x, y, z)\) on the ball \(B = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1\}\), we would have to find a parameterization of the sphere \(S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}\). This more difficult, and in any case only leads to a two-variable problem.

**Method 2: Lagrange Multipliers.**

The basic idea is the following. If a point \((x, y)\) on \(T\) is a local maximum or local minimum for \(f\), then the gradient of \(f\) does not need to equal the zero vector, but is does need to be perpendicular to the circle (for otherwise, the function would increase along the circle in the direction in which \(\nabla f(x, y)\) is “increasing". On the other hand, if we let \(g(x, y) = x^2 + y^2\), then the circle \(T\) is the level set of \(g\) passing through the point \((x, y)\), and we know that the gradient of \(g\), \(\nabla g(x, y)\), is perpendicular to the level set. Thus \(\nabla f(x, y)\) and \(\nabla g(x, y)\) point in the same direction, and hence there must be a real number \(\lambda\) so that \(\nabla f(x, y) = \lambda \nabla g(x, y)\).

We think of \(\lambda\) as an additional unknown (it is called the Lagrange multiplier), and we are thus led to solving three equations in three unknowns \((x, y, \lambda)\):

\[
\frac{\partial f}{\partial x}(x, y) = \lambda \frac{\partial g}{\partial x}(x, y),
\]

\[
\frac{\partial f}{\partial y}(x, y) = \lambda \frac{\partial g}{\partial y}(x, y),
\]

\[
g(x, y) = 1.
\]

In this example, this amounts to

\[
2xy = 2\lambda x,
\]

\[
x^2 = 2\lambda y,
\]

\[
x^2 + y^2 = 1.
\]

If \(x = 0\), we get \(y = \pm 1\) from the third equation, and this gives us two candidates \((0, 1)\) and \((0, -1)\). If \(x \neq 0\), it follows from the first equation that \(y = \lambda\), and then from the second equation that \(x^2 = 2\lambda y^2\). Plugging this into the third equation gives \(3\lambda^2 = 1\), and so we recover the various solutions that we found with the parameterization method.

**Example 2:** Find the shortest distance from the origin to the surface

\[
g(x, y, z) = x^2 - xy + y^2 - z^2 = 1.
\]
We are trying to minimize the function \( f(x, y, z) = x^2 + y^2 + z^2 \) subject to the constraint \( g(x, y, z) = 1 \). At a point where the distance is minimum, we must have the gradients of \( f \) and \( g \) pointing in the same direction. Thus we need to solve
\[
\frac{\partial f}{\partial x}(x, y, z) = \lambda \frac{\partial g}{\partial x}(x, y, z),
\frac{\partial f}{\partial y}(x, y, z) = \lambda \frac{\partial g}{\partial y}(x, y, z),
\frac{\partial f}{\partial z}(x, y, z) = \lambda \frac{\partial g}{\partial z}(x, y, z),
g(x, y, z) = 1.
\]
This means we have to solve
\[
2x = \lambda(2x - y),
2y = \lambda(-x + 2y),
2z = \lambda(-2z),
x^2 - xy + y^2 - z^2 = 1.
\]
If \( z \neq 0 \) it follows from the third equation that \( \lambda = -1 \), and the first two equations become
\[
4x - y = 0,
x + 4y = 0,
\]
from which it follows that \( x = y = 0 \). However, this is incompatible with the fourth equation. Thus we must have \( z = 0 \), and we need to solve
\[
(2 - 2\lambda)x + \lambda y = 0,
\lambda x + (2 - 2\lambda)y = 0,
x^2 - xy + y^2 = 1.
\]

**Example 3:** Let \( C \) be the curve of intersection of the two surfaces
\[
g_1(x, y, z) = x^2 - xy + y^2 - z^2 = 1,
g_2(x, y, z) = x^2 + y^2 = 1.
\]
Find the point on \( C \) which is closest to the origin.

We are trying to minimize the function \( f(x, y, z) = x^2 + y^2 + z^2 \) subject to the two constraints \( g_1(x, y, z) = 1 \) and \( g_2(x, y, z) = 1 \). At the minimum point \( a \), the gradient \( \nabla f(a) \) must be perpendicular to the curve \( C \). However, both \( \nabla g_1(a) \) and \( \nabla g_2(a) \) are also perpendicular to \( C \), so we should solve the system of four equations in four unknowns:
\[
\frac{\partial f}{\partial x}(x, y, z) = \lambda_1 \frac{\partial g_1}{\partial x}(x, y, z) + \lambda_1 \frac{\partial g_2}{\partial x}(x, y, z),
\frac{\partial f}{\partial y}(x, y, z) = \lambda_1 \frac{\partial g_1}{\partial y}(x, y, z) + \lambda_1 \frac{\partial g_2}{\partial y}(x, y, z),
\frac{\partial f}{\partial z}(x, y, z) = \lambda_1 \frac{\partial g_1}{\partial z}(x, y, z) + \lambda_1 \frac{\partial g_2}{\partial z}(x, y, z),
g_1(x, y, z) = 1,
g_2(x, y, z) = 1.
\]
This means we have to solve
\[
2x = \lambda_1(2x - y) + \lambda_2(2x),
2y = \lambda_1(-x + 2y) + \lambda_2(2y),
2z = \lambda_1(-2z),
x^2 - xy + y^2 - z^2 = 1,
x^2 + y^2 = 1.
\]
Example 3: Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a symmetric linear mapping. This means that for all $x, y \in \mathbb{R}^n$ we have $\langle Tx, y \rangle = \langle x, Ty \rangle$. Show that there is a real number $\lambda$ so that if the function $g(x) = \langle Tx, x \rangle$ attains its maximum (or minimum) subject to the constraint $||x|| = 1$ at a point $a$, then $T[a] = \lambda a$. 