9.3 - The Chain Rule

9.3a - Lemma: Let $U \subset \mathbb{R}^p$ be open, let $a \in U$, let $r > 0$ s.t. $B(a) \subset U$, let $A$ be a $p \times r$ matrix and let $F: U \rightarrow \mathbb{R}^q$ be a map. Then $F$ is diff. at $a$ with $A = [dF(a)]$ iff there is a map $E: B(a) \rightarrow \mathbb{R}^q$ s.t.

1) $E(a + h) = F(a) + Ah + E(h)\|h\| \text{ for all } h \in \mathbb{R}^p \text{ s.t. } 0 < \|h\| < r,$

2) $\lim_{h \to 0} E(h) = 0.$

9.3b - Notes:

1) Students are asked to prove 9.3a in HW 9.5.

2) We may assume $E$ is chs at 0 by defining $E(0) = 0.$

9.3c - Theorem (Chain Rule): Let $U \subset \mathbb{R}^p, V \subset \mathbb{R}^q$ be open sets, let $G: U \rightarrow \mathbb{R}^q, F: V \rightarrow \mathbb{R}^r$ be maps and let at $U$ s.t. $b = G(a) \in V.$ Sps that $G$ is dif. at $a$ and that $F$ is dif. at $b.$ Then $F \circ G$ is dif. at $a$ and $d(F \circ G)(a) = dF(b) \circ dG(a).$

1) Set $A = [dG(a)]$ and $B = [dF(b)].$ Since $G$ is dif. at $a$ and $F$ is dif. at $b,$ by 9.3a, there exist for $\varepsilon, \eta$ defined on neighborhoods of 0 of $F$ and $G$ at $0$ with $E(0) = 0$ and $\eta(0) = 0,$ s.t.

\[ G(a + h) = G(a) + Ah + \varepsilon(h)\|h\| \text{ for } h \text{ in a } \text{nbd of } 0, \text{ and } \]

\[ F(b + k) = F(b) + Bk + \eta(k)\|k\| \text{ for } k \text{ in a } \text{nbd of } 0. \]

Then with $e = G(a + h) - G(a) = Ah + \varepsilon(h)\|h\|$ we have $b + k = G(a + h)$ and, if $h$ is chosen s.t. $\|h\|$ is sufficiently small,

\[ (F \circ G)(a + h) - (F \circ G)(a) = F(b + k) - F(b) = Bk + \eta(k)\|k\| \]

\[ = \|B\| \|k\| \|k\| + \eta(k)\|k\| \leq \|B\| \|E(h)\| \|h\| + \|\eta(k)\|\|k\| \leq \|B\| \varepsilon(h)\|h\| + \|\eta(k)\|\|k\|. \]

So

\[ \|F \circ G\| = \|F\| \|G\| \|h\| \]

\[ \leq \|B\| \|E(h)\| \|h\| + \|\eta(k)\|\|k\| \leq \|B\| \varepsilon(h)\|h\| + \|\eta(k)\|\|k\|. \]

Then since

\[ \lim_{h \to 0} h = \lim_{h \to 0} \varepsilon(h)\|h\| = 0, \text{ we have } \lim_{h \to 0} \eta(k)\|k\| = 0 \text{ and so } \]

\[ \lim_{h \to 0} \|B\| \varepsilon(h)\|h\| + \|\eta(k)\|\|k\| = 0. \]

Therefore,

\[ \lim_{h \to 0} \frac{1}{\|h\|} [F \circ G(a + h) - F \circ G(a) - B\|A\|h] = 0. \text{ Hence } F \circ G \text{ is dif. at } a \text{ and } \]

\[ [d(F \circ G)(a)] = BA = [dF(b)] [dG(a)] \text{ or equivalently } \]

\[ d(F \circ G)(a) = dF(b) \circ dG(a). \]

9.3d - Ex: Let $U \subset \mathbb{R}^d$ be open and let $f, g: U \rightarrow \mathbb{R}$ be maps. Show that if $f$ and $g$ are both diff. at $a$, then $g 

9.3e - Ex: Let $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ be a $C^1$ function $\mathbb{R}^3$ and define $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ by $g(x, y) = f(x, y, z)$ for all $(x, y, z) \in \mathbb{R}^3$ and define $H: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ by $H(x, y, z) = (y, z)$ for all $(x, y, z) \in \mathbb{R}^3$. Show that $g$ is $C^1$ on $\mathbb{R}^3$ and find $[dg(1, 2, 3)].$

1) By straightforward computation we have

\[ \frac{\partial g}{\partial x} = 0, \frac{\partial g}{\partial y} = z, \frac{\partial g}{\partial z} = y, \frac{\partial H}{\partial x} = z, \frac{\partial H}{\partial y} = 0, \frac{\partial H}{\partial z} = 0. \]

Since these are all chs, $H$ is $C^1$ on $\mathbb{R}^3$. Observe that $g = f \circ H.$
9.3 - continued

For all \((x,y,z) \in \mathbb{R}^3\), \(f\) is diff at \((x,y,z)\) and \(f\) is diff at \(H(x,y,z)\), so by the chain rule \(g = f \circ H\) is diff at \((x,y,z)\) and

\[
\frac{dg(x,y,z)}{dx} = \left[ \frac{df}{dH} \right] \left[ \frac{dH(x,y,z)}{dx} \right]
\]

\[
= \left( \frac{\partial f}{\partial x}(H(x,y,z)), \frac{\partial f}{\partial y}(H(x,y,z)), \frac{\partial f}{\partial z}(H(x,y,z)) \right) \begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

So we have:

\[
\frac{\partial g}{\partial x} = 2 \frac{\partial f}{\partial x}(H(x,y,z)) + y \frac{\partial f}{\partial y}(H(x,y,z)), \quad \frac{\partial g}{\partial y} = 2 \frac{\partial f}{\partial y}(H(x,y,z)) + z \frac{\partial f}{\partial z}(H(x,y,z))\]

and

\[
\frac{\partial g}{\partial z} = 2 \frac{\partial f}{\partial z}(H(x,y,z)) + x \frac{\partial f}{\partial x}(H(x,y,z)).
\]

Since all first partials are cts, \(g\) is \(C^1\) on \(\mathbb{R}^3\).

If \(u = (x,y,z) = (1,2,3)\) we have \(b = H(1,2,3) = (6,3,2)\) and so

\[
\frac{dg(u)}{dx} = \left[ \frac{df(b)}{dH(u)} \right] \left[ \frac{dH(u)}{dx} \right] = \begin{pmatrix}
3/2 & 2/3 & 1/2 \\
3/2 & 2/3 & 1/2 \\
0 & 0 & 1
\end{pmatrix}
\]

\[
= \begin{pmatrix}
3/2 & 2/3 & 1/2 \\
3/2 & 2/3 & 1/2 \\
0 & 0 & 1
\end{pmatrix}
\]

9.3f Note: Let \(I \subset \mathbb{R}\) be an open interval and let \(g: I \rightarrow \mathbb{R}^d\) be a function; write \(g(t) = (x_1(t), \ldots, x_d(t))\). Then \(g\) is diff at \(t \in I\) if \(g'(t) = (x'_1(t), \ldots, x'_d(t))\) exist for \((t, \ldots, t)\). In this case, \(g'(t) = \begin{bmatrix} x'_1(t) \\ \vdots \\ x'_d(t) \end{bmatrix}\). Denote this by \(g'(t)\).

9.3g Cor: Let \(I \subset \mathbb{R}\) be an open interval and let \(\varphi: I \rightarrow \mathbb{R}^d\) be a function which is diff at \(t_0\). Let \(U \subset \mathbb{R}^d\) be open, \(\varphi(\alpha) = \beta \in U\) and let \(h: U \rightarrow \mathbb{R}\) be a function which is diff at \(\alpha\). Then \(h \circ \varphi\) is diff at \(t_0\) and

\[
(h \circ \varphi)'(t_0) = \left[ \frac{dh(\varphi(t))}{dx} \right] \left[ \frac{d\varphi(t)}{dx} \right] = \frac{dh}{dx}(\varphi(t_0)) \sum_{i=1}^d \frac{\partial \varphi_i}{\partial x_i}(\varphi(t_0)) \frac{d\varphi_i}{dx}(t_0).
\]

This follows easily from the chain rule (Theorem 9.3c). This is the usual version of the chain rule taught in Multivariable Calculus.

9.3h Notes: Curves in \(\mathbb{R}^d\) are parametrized by functions of the form \(r: I \rightarrow \mathbb{R}^d\) where \(I \subset \mathbb{R}\) is an interval. Any line in \(\mathbb{R}^d\) is determined by a unit vector \(u \subset \mathbb{R}^d\), and a pt on \(\mathbb{R}^d\) via the formula \(a(t) = at + u\) for \(t \in \mathbb{R}\). Moreover \(r\) is diff on \(I\) if and only if \(r'(t) = u\) for all \(t \in I\).

Now supp that \(U \subset \mathbb{R}^d\) is open s.t. \(a \in U\) and supp that \(h: U \rightarrow \mathbb{R}\) is diff at \(a\). Then given a unit vec. \(u \subset \mathbb{R}^d\), by Cor. 9.3g, the directional derivative of \(h\) at \(a\) in the direction of \(u\) exists and is given by

\[
D_u h(a) = dh(a)u = \sum_{i=1}^d \frac{dh}{dx_i}(a) u_i.
\]

9.3i Notes: Let \(U \subset \mathbb{R}^d\) be open and let \(F: U \rightarrow \mathbb{R}^k\) be a \(C^1\) map on \(U\). Let \(M_{q,p}\) denote the vector space of \(q \times p\) matrices equipped with the matrix norm \(\|M\| = \max_{\|x\|=1} \|Mx\|\). Then it can be shown that the map \(x \in U \mapsto [DF(x)] \in M_{q,p}\) is cts. Moreover, with notation as in 9.3c, \(f \circ g\) is \(C^1\) on \(U\) and \(F\) is \(C^1\) on \(V\), then \(F \circ g\) is \(C^1\) on \(g^{-1}(V)\).