Chain Rule

This document will derive the chain rule for composition of differentiable mappings. First a definition.

Definition 1. *Suppose $\mathbb{R}^m \xrightarrow{F} \mathbb{R}^n$ is defined in an open set around a. We say that F is differentiable at a if*

$$F(x) = F(a) + P(x)(x-a),$$

where P is continuous at a. In this notation P is an $n \times m$ function-valued matrix and $x - a$ is an $m \times 1$ column vector. We define the derivative *$DF(a)$ of F at a to be the value $DF(a) = P(a)$. It's not hard to verify that if we let $F(x) = [f_1, \ldots, f_n]^T$, then*

$$\begin{bmatrix} (f_1)_{x_1} & (f_1)_{x_2} & \ldots & (f_1)_{x_m} \\ (f_2)_{x_1} & & \ldots & (f_2)_{x_m} \\ \ldots & & & \ldots \\ (f_n)_{x_1} & (f_n)_{x_2} & \ldots & (f_n)_{x_m} \end{bmatrix} (a).$$

Theorem 1. *Suppose $\mathbb{R}^k \xrightarrow{G} \mathbb{R}^m \xrightarrow{F} \mathbb{R}^n$ are such that G is differentiable at a and F is differentiable at $b = G(a)$. Then $\mathbb{R}^k \xrightarrow{H} \mathbb{R}^n$ defined by $H(x) = F(G(x))$ is differentiable at a and*

$$DH(a) = DF(b)DG(a),$$

(matrix multiply).

Proof.

$$\begin{aligned} H(a+x) &= F(G(a+x)) = F(G(a) + Q(x)(x-a)) \\ &= F(G(a)) + P(G(a) + Q(x)(x-a))[Q(x)(x-a)] \\ &= F(b) + P(b + Q(x)(x-a))[Q(x)(x-a)]. \end{aligned}$$

Since $Q(x)$ is continuous at a and P is continuous at $b = G(a)$, $P(G(a) + Q(x)(x-a))Q(x)$ is continuous at $x = a$. This proves differentiability. The derivative is

$$P(G(a))Q(a) = P(b)Q(a) = DF(b)Dg(a).$$

□