This note will give a proof for the higher dimensional version of the chain rule. It will make use of the following simple fact:

Lemma. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map. Then, there is a constant $C \geq 0$ such that*

$$\|T(\mathbf{x})\| \leq C\,\|\mathbf{x}\|$$

for every $\mathbf{x} \in \mathbb{R}^n$.

Proof. Let $A = [a_{ij}]$ be the $m \times n$ matrix representing T, so $T(x_1, \ldots x_n) = (y_1, \ldots, y_m)$, where

$$y_i = \sum_{j=1}^{n} a_{ij}\, x_j \tag{1}$$

for every $1 \leq i \leq m$. Let

$$\alpha = \max_{i,j} |a_{ij}| \qquad \text{and} \qquad \ell(\mathbf{x}) = \max_{j} |x_j|.$$

Then

$$\|\mathbf{x}\| = (x_1^2 + \cdots + x_n^2)^{1/2} \geq \ell(\mathbf{x}).$$

By (1) and the triangle inequality,

$$|y_i| = \left|\sum_{j=1}^{n} a_{ij} x_j\right| \leq \sum_{j=1}^{n} |a_{ij}|\,|x_j| \leq n\,\alpha\,\ell(\mathbf{x}),$$

which shows

$$\|T(\mathbf{x})\| = (y_1^2 + \cdots + y_m^2)^{1/2} \leq \sqrt{m}\, n\,\alpha\,\ell(\mathbf{x}) \leq \sqrt{m}\, n\,\alpha\,\|\mathbf{x}\|.$$

Thus, the lemma holds if we set $C = \sqrt{m}\; n\; \alpha$. □

Theorem (The Chain Rule). *Suppose $g : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{a}$ and $f : \mathbb{R}^m \to \mathbb{R}^k$ is differentiable at $\mathbf{b} = g(\mathbf{a})$. Then, the composition $h = f \circ g : \mathbb{R}^n \to \mathbb{R}^k$ is differentiable at $\mathbf{a}$ and*

$$Dh(\mathbf{a}) = Df(\mathbf{b})\, Dg(\mathbf{a}).$$

Proof. To simplify the notation, denote $Dg(\mathbf{a})$ by T and $Df(\mathbf{b})$ by L. Thus, differentiability of g at $\mathbf{a}$ means we can write

$$g(\mathbf{a} + \mathbf{v}) = g(\mathbf{a}) + T(\mathbf{v}) + \|\mathbf{v}\|\; E_g, \tag{2}$$

where $\lim_{\mathbf{v}\to\mathbf{0}} \|E_g\| = 0$. Similarly, differentiability of f at $\mathbf{b}$ means we can write

$$f(\mathbf{b} + \mathbf{u}) = f(\mathbf{b}) + L(\mathbf{u}) + \|\mathbf{u}\|\; E_f, \tag{3}$$

where $\lim_{\mathbf{u}\to\mathbf{0}} \|E_f\| = 0$.

For each $\mathbf{v} \neq \mathbf{0}$, set

$$\mathbf{u} = g(\mathbf{a} + \mathbf{v}) - \mathbf{b} = T(\mathbf{v}) + \|\mathbf{v}\| \, E_g.$$

Continuity of g at $\mathbf{a}$ shows that $\mathbf{u} \to \mathbf{0}$ whenever $\mathbf{v} \to \mathbf{0}$. By (2), (3), and linearity of L,

$$\begin{aligned}
h(\mathbf{a} + \mathbf{v}) &= f(g(\mathbf{a} + \mathbf{v})) = f(\mathbf{b} + \mathbf{u}) \\
&= f(\mathbf{b}) + L(\mathbf{u}) + \|\mathbf{u}\| \, E_f \\
&= h(\mathbf{a}) + L(T(\mathbf{v}) + \|\mathbf{v}\| \, E_g) + \|\mathbf{u}\| \, E_f \\
&= h(\mathbf{a}) + LT(\mathbf{v}) + \|\mathbf{v}\| \, L(E_g) + \|\mathbf{u}\| \, E_f \\
&= h(\mathbf{a}) + LT(\mathbf{v}) + \|\mathbf{v}\| \left(L(E_g) + \frac{\|\mathbf{u}\|}{\|\mathbf{v}\|} \, E_f \right).
\end{aligned}$$

Thus, the claim $Dh(\mathbf{a}) = LT$ will be proved once we show that

$$\lim_{\mathbf{v} \to \mathbf{0}} L(E_g) + \frac{\|\mathbf{u}\|}{\|\mathbf{v}\|} \, E_f = 0. \tag{4}$$

To see this, note that by the Lemma, there is a constant $C_1 \geq 0$ such that

$$\|T(\mathbf{x})\| \leq C_1 \|\mathbf{x}\|$$

for every $\mathbf{x} \in \mathbb{R}^n$. Similarly, there is a constant $C_2 \geq 0$ such that

$$\|L(\mathbf{y})\| \leq C_2 \|\mathbf{y}\|$$

for every $\mathbf{y} \in \mathbb{R}^m$. Hence,

$$\begin{aligned}
\left\| L(E_g) + \frac{\|\mathbf{u}\|}{\|\mathbf{v}\|} \, E_f \right\| &\leq \|L(E_g)\| + \frac{\|\mathbf{u}\|}{\|\mathbf{v}\|} \|E_f\| \\
&\leq C_2 \|E_g\| + \frac{\|T(\mathbf{v}) + \|\mathbf{v}\| E_g\|}{\|\mathbf{v}\|} \|E_f\| \\
&\leq C_2 \|E_g\| + \frac{\|T(\mathbf{v})\| + \|\mathbf{v}\| \|E_g\|}{\|\mathbf{v}\|} \|E_f\| \\
&\leq C_2 \|E_g\| + (C_1 + \|E_g\|) \|E_f\|.
\end{aligned}$$

As $\mathbf{v} \to \mathbf{0}$, both $\|E_g\|$ and $\|E_f\|$ tend to zero, which proves (4).