Chain Rule Derivation

Let γ be a curve in $\mathbb{R}^n$ given by the differentiable vector-valued function $\mathbf{v}(t)$ where $\mathbf{v}$ has n components v^μ, each a differentiable function of t. Let $f(\mathbf{x})$ be a differentiable scalar function on $\mathbb{R}^n$, and let $f(\mathbf{v}(t))$ be the composition of functions $f(\mathbf{x})$ and $\mathbf{v}(t)$, *i.e.* f evaluated on γ. We want to find the derivative of f as we move along γ:

$$\frac{\mathrm{d}f}{\mathrm{d}t} \coloneqq \lim_{\Delta t \to 0} \frac{f(\mathbf{v}(t_0 + \Delta t)) - f(\mathbf{v}_0)}{\Delta t}$$

where $\mathbf{v}_0 \coloneqq \mathbf{v}(t_0)$.

Let's begin by defining the vector $\Delta \mathbf{v} \coloneqq \mathbf{v}(t_0 + \Delta t) - \mathbf{v}_0 \equiv \Delta v^\mu \hat{x}_\mu$. Now let's imagine the graph of f fully represented in $\mathbb{R}^{n+1}$. We can imagine the tangent space to this graph at any given point $\mathbf{x}_0$ as a hyperplane defined by the graph of

$$g_{\mathbf{x}_0}(\mathbf{x}) \coloneqq \frac{\partial f}{\partial x^\mu}(x^\mu - x_0^\mu) + f(\mathbf{x}_0)$$

Another way to say this is that f can be linearly approximated by the first two terms of its Taylor expansion (which is all that g is). Let's now define $\Delta f \coloneqq f(\mathbf{v}(t_0 + \Delta t)) - f(\mathbf{v}_0)$. Δf can be linearly approximated using $g_{\mathbf{v}_0}$:

$$\begin{aligned}
\Delta f &\approx g_{\mathbf{v}_0}(\mathbf{v}(t_0 + \Delta t)) - g_{\mathbf{v}_0}(\mathbf{v}_0) \\
&\approx \frac{\partial f}{\partial x^\mu}(v^\mu(t_0 + \Delta t) - v_0^\mu) + f(\mathbf{v}_0) - f(\mathbf{v}_0) \\
&\approx \frac{\partial f}{\partial x^\mu}\Delta v^\mu
\end{aligned}$$

We can turn this approximation into an exact equality simply by introducing some additional functions:

$$\Delta f = \frac{\partial f}{\partial x^\mu}\Delta v^\mu + \epsilon_\mu \Delta v^\mu$$

where each ϵ is some corrective function that approaches zero as Δv^μ approaches zero. This should make intuitive sense because as Δv^μ approaches zero, the relevant patch of the graph of f is increasingly better approximated by the tangent hyperplane. Let us now divide this equation by Δt:

$$\frac{\Delta f}{\Delta t} = \frac{\partial f}{\partial x^\mu}\frac{\Delta v^\mu}{\Delta t} + \epsilon_\mu \frac{\Delta v^\mu}{\Delta t}$$

With how we defined Δf, it should be clear that taking the limit here as $\Delta t \to 0$ yields the derivative we were looking for. It should also be noted that as $\Delta t \to 0$, $\Delta v^\mu \to 0$ as well, which means that the ϵ_μ functions vanish. So we are left with

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial x^\mu}\frac{\mathrm{d}v^\mu}{\mathrm{d}t}$$