Then since $b = f(a)$, using the above we have

$$\begin{aligned}(g \circ f)(a+s) &= g(f(a+s)) = g(b + \mathrm{D}f_a(s) + \epsilon_1(s)) \\ &= g(b) + \mathrm{D}g_b(\mathrm{D}f_a(s) + \epsilon_1(s)) + \epsilon_2(\mathrm{D}f_a(s) + \epsilon_1(s)) \\ &= g(b) + (\mathrm{D}g_b \circ \mathrm{D}f_a)(s) + \mathrm{D}g_b(\epsilon_1(s)) + \epsilon_2(\mathrm{D}f_a(s) + \epsilon_1(s)).\end{aligned}$$

Now by $(*_1)$ and $(*_2)$ we have

$$\begin{aligned}\|\mathrm{D}g_b(\epsilon_1(s)) + \epsilon_2(\mathrm{D}f_a(s) + \epsilon_1(s))\| &\leq \|\mathrm{D}g_b(\epsilon_1(s))\| + \|\epsilon_2(\mathrm{D}f_a(s) + \epsilon_1(s))\| \\ &\leq \eta \|\mathrm{D}g_b\| \|s\| + \eta(\|\mathrm{D}f_a\| + 1) \|s\| \\ &= \eta(\|\mathrm{D}f_a\| + \|\mathrm{D}g_b\| + 1) \|s\|,\end{aligned}$$

so if we write $\epsilon_3(s) = \mathrm{D}g_b(\epsilon_1(s)) + \epsilon_2(\mathrm{D}f_a(s) + \epsilon_1(s))$ we proved that

$$(g \circ f)(a+s) = g(b) + (\mathrm{D}g_b \circ \mathrm{D}f_a)(s) + \epsilon_3(s)$$

with $\epsilon_3(s) \leq \eta(\|\mathrm{D}f_a\| + \|\mathrm{D}g_b\| + 1) \|s\|$, which proves that $\mathrm{D}g_b \circ \mathrm{D}f_a$ is the derivative of $g \circ f$ at a. Since $\mathrm{D}f_a$ and $\mathrm{D}g_b$ are continuous, so is $\mathrm{D}g_b \circ \mathrm{D}f_a$, which proves our proposition. ☐

Theorem 39.6 has many interesting consequences. We mention two corollaries.

Proposition 39.7. *Given three normed affine spaces E, F, and G, for any open subset A in E, for any $a \in A$, let $f \colon A \to F$ such that $\mathrm{D}f(a)$ exists, and let $g \colon F \to G$ be a continuous affine map. Then, $\mathrm{D}(g \circ f)(a)$ exists, and*

$$\mathrm{D}(g \circ f)(a) = \overrightarrow{g} \circ \mathrm{D}f(a),$$

where $\overrightarrow{g}$ is the linear map associated with the affine map g.

Proposition 39.8. *Given two normed affine spaces E and F, let A be some open subset in E, let B be some open subset in F, let $f \colon A \to B$ be a bijection from A to B, and assume that $\mathrm{D}f$ exists on A and that $\mathrm{D}f^{-1}$ exists on B. Then, for every $a \in A$,*

$$\mathrm{D}f^{-1}(f(a)) = (\mathrm{D}f(a))^{-1}.$$

Proposition 39.8 has the remarkable consequence that the two vector spaces $\overrightarrow{E}$ and $\overrightarrow{F}$ have the same dimension. In other words, a local property, the existence of a bijection f between an open set A of E and an open set B of F, such that f is differentiable on A and f^{-1} is differentiable on B, implies a global property, that the two vector spaces $\overrightarrow{E}$ and $\overrightarrow{F}$ have the same dimension.

Let us mention two more rules about derivatives that are used all the time.

Let $\iota \colon \mathbf{GL}(n, \mathbb{C}) \to \mathrm{M}_n(\mathbb{C})$ be the function (inversion) defined on invertible $n \times n$ matrices by

$$\iota(A) = A^{-1}.$$