Theorem. Let $X_1, \ldots, X_k, Y$ be Banach spaces, let $D_1 \subset X_1, \ldots, D_k \subset X_k$ be open sets, let $f : D_1 \times \cdots \times D_k \to Y$ be a function. Assume that at each point $(x_1, \ldots, x_k) \in D_1 \times \cdots \times D_2$, each partial derivative $D_i f(x_1, \ldots, x_k) \in L(X_i, Y)$ exists, and each map $D_i f : D_1 \times \cdots \times D_k \to L(X_i, Y)$ is continuous. Then f is continuously differentiable, and

$$Df(x_1, \ldots, x_k)(h_1, \ldots, h_k) = D_1 f(x_1, \ldots, x_k) h_1 + \cdots + D_k f(x_1, \ldots, x_k) h_k.$$

Proof for the case $k = 2$. The following calculation shows that $Df(x_1, x_2)$ is the bounded linear map $(h_1, h_2) \to D_1 f(x_1, x_2) h_1 + D_2 f(x_1, x_2) h_2$:

$$f(x_1+h_1, x_2+h_2) - f(x_1, x_2) - (D_1 f(x_1, x_2) h_1 + D_2 f(x_1, x_2) h_2)$$

$$= f(x_1+h_1, x_2+h_2) - f(x_1+h_1, x_2) + f(x_1+h_1, x_2) - f(x_1, x_2)$$

$$- (D_1 f(x_1, x_2) h_1 + D_2 f(x_1, x_2) h_2)$$

$$= \left(\int_0^1 D_2 f(x_1+th_1, x_2+th_2) - D_2 f(x_1, x_2)\, dt\right) h_2$$

$$+ \left(\int_0^1 D_1 f(x_1+th_1, x_2) - D_1 f(x_1, x_2)\, dt\right) h_1 .$$

Therefore, given $\varepsilon > 0$,

$$\| f(x_1+h_1, x_2+h_2) - f(x_1,x_2) - (D_1 f(x_1,x_2) h_1 + D_2 f(x_1,x_2) h_2) \| \leq$$

$$\sup_{0 \leq t \leq 1} \| D_2 f(x_1+h_1, x_2+th_2) - D_2 f(x_1,x_2) \| \|h_2\| +$$

$$\sup_{0 \leq t \leq 1} \| D_1 f(x_1+th_1, x_2) - D_1 f(x_1,x_2) \| \|h_1\| \leq \frac{\varepsilon}{2}\|h_2\| + \frac{\varepsilon}{2}\|h_1\|$$

$\leq \varepsilon \| (h_1,h_2) \|$ for $\| (h_1,h_2) \|$ sufficiently small, because

$D_1 f$ and $D_2 f$ depend continuously on (x_1,x_2).

To show that $Df(x_1,x_2)$ depends continuously on (x_1,x_2):
Fix (x_1,x_2).

$$\left(Df(x_1,x_2) - Df(x_1',x_2')\right)(h_1,h_2) =$$

$$(D_1 f(x_1,x_2) - D_1 f(x_1',x_2'))h_1 + (D_2 f(x_1,x_2) - D_2 f(x_1',x_2'))h_2$$

Therefore $\| (Df(x_1,x_2) - Df(x_1',x_2'))(h_1,h_2) \| \leq$
$\frac{\varepsilon}{2}\|h_1\| + \frac{\varepsilon}{2}\|h_2\| \leq \varepsilon \| (h_1,h_2) \|$ for $\| (x_1,x_2) - (x_1',x_2') \|$ sufficiently small, again because $D_1 f$ and $D_2 f$ depend continuously on (x_1,x_2). Hence for $\| (x_1,x_2) - (x_1',x_2') \|$ sufficiently small, $\| Df(x_1,x_2) - Df(x_1',x_2') \| \leq \varepsilon$.