Theorem: *If a function $f(x)$ is defined for x in a ball $B(x_o, R) \subseteq \mathbb{R}^n$ and if the partial derivatives $\frac{\partial f}{\partial x_j}$, for $1 \le j \le n$, are all defined in $B(x_o, R)$ and if each partial derivative $\frac{\partial f}{\partial x_j}$ is continuous at x_o, then f is differentiable at x_o.*

Proof: We have to show that $\frac{|\text{Rem}|}{|x-x_o|} \to 0$ as $x \to x_o$, where Rem stands for the Remainder after approximating $f(x)$ by the "linear" function $f(x_o) + \nabla f(x_o) \bullet (x - x_o)$:

$$\text{Rem} = f(x) - f(x_o) - \nabla f(x_o) \bullet (x - x_o).$$

We will use the MVT to work partial derivatives into the part $f(x) - f(x_o)$ of the Remainder. We begin by using Subtracting-and-Adding n times to change one coordinate at a time to get from x to x. We set

$$
\begin{aligned}
&w_0 = x = (x_1, x_2, x_3, \ldots, x_n) \ \text{(no coordinates changed)},\\
&w_1 = (x_{o1}, x_2, x_3, \ldots, x_n) \ \text{(first coordinate changed)},\\
&w_2 = (x_{o1}, x_{o2}, x_3, \ldots, x_n) \ \text{(first two coordinates changed)},\\
&w_3 = (x_{o1}, x_{o2}, x_{o3}, \ldots, x_n) \ \text{(first three coordinates changed)},\\
&\qquad\qquad\qquad\qquad \ldots\ldots\ldots\ldots\ldots\\
&\qquad\qquad\qquad\qquad \ldots\ldots\ldots\ldots\ldots\\
&\qquad\qquad\qquad\qquad \ldots\ldots\ldots\ldots\ldots\\
&w_k = (x_{o1}, x_{o2}, x_{o3}, \ldots, x_{o,k-1},, x_{o,k-1}, x_{k+1}, \ldots, x_n) \ \text{(first } k \text{ coordinates changed)},\\
&\qquad\qquad\qquad\qquad \ldots\ldots\ldots\ldots\ldots\\
&\qquad\qquad\qquad\qquad \ldots\ldots\ldots\ldots\ldots\\
&\qquad\qquad\qquad\qquad \ldots\ldots\ldots\ldots\ldots\\
&w_n = x_o = (x_{o1}, x_{o2}, x_{o3}, \ldots, x_{on}) \ \text{(all coordinates changed)}.
\end{aligned}
$$

Then

$$
\begin{aligned}
f(x) - f(x_o) &= f(w_0) - f(w_n)\\
&= f(w_0) - f(w_1)\\
&+ f(w_1) - f(w_2)\\
&+ f(w_2) - f(w_3)\\
&+ \cdots\cdots\cdots +\\
&+ f(w_{k-1}) - f(w_k)\\
&+ f(w_k) - f(w_{k+1})\\
&+ \cdots\cdots\cdots +\\
&+ f(w_{n-1}) - f(w_n)\\
&= \sum_{k=1}^{n} (f(w_{k-1}) - f(w_k)).
\end{aligned}
\tag{1}
$$

The k-th coordinate of w_{k-1} is x_k and the k-th coordinate of w_k is x_{ok}. In all the other coordinates of these two vectors, they agree. Thus we can use the MVT on $f(w_{k-1}) - f(w_k)$ by considering the function

$$g(t) := f(x_{o1}, \ldots, x_{o,k-1}, t, x_{k+1}, \ldots, x_n)$$

of the one variable t. According to MVT, $g(x_k) - g(x_{ok}) = g'(c_k)(x_k - x_{ok})$ for some c_k that lies strictly between x_{ok} and x_k. But by the way we defined $g(t)$, $g(x_k) - g(x_{ok}) = f(w_{k-1}) - f(w_k) = g'(c_k)(x_k - x_{ok})$. What is $g'(c_k)$, in terms of f and its various partials? The answer is:

$$g'(t) = \frac{\partial f}{\partial x_k}(x_{o1}, \ldots, x_{o,k-1}, t, x_{k+1}, \ldots, x_n), \quad \text{so that}$$

$$g'(c_k) = \frac{\partial f}{\partial x_k}(x_{o1},\, \ldots,\, x_{o,k-1},\, c_k,\, x_{k+1},\, \ldots,\, x_n).$$

Let's now define $w_k^* := (x_{o1},\, \ldots,\, x_{o,k-1},\, c_k,\, x_{k+1},\, \ldots,\, x_n)$. This allows us to write $f(w_{k-1}) - f(w_k) = g'(c_k)(x_k - x_{ok}) = \frac{\partial f}{\partial x_k}(w_k^*)(x_k - x_{ok})$. Thus we can re-write (1) *with* partial derivatives worked in, as

$$\begin{aligned} f(x) - f(x_o) &= f(w_0) - f(w_n) \\ &= \frac{\partial f}{\partial x_1}(w_1^*)(x_1 - x_{o1}) \\ &+ \frac{\partial f}{\partial x_2}(w_2^*)(x_2 - x_{o2}) \\ &+ \frac{\partial f}{\partial x_3}(w_3^*)(x_3 - x_{o3}) \\ &+ \cdots\cdots\cdots + \\ &+ \frac{\partial f}{\partial x_{k-1}}(w_{k-1}^*)(x_{k-1} - x_{o,k-1}) \\ &+ \frac{\partial f}{\partial x_k}(w_k^*)(x_k - x_{ok}) \\ &+ \cdots\cdots\cdots + \\ &+ \frac{\partial f}{\partial x_n}(w_n^*)(x_n - x_{on}) \\ &= \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(w_k^*)(x_k - x_{ok}). \end{aligned} \tag{2}$$

This lets us write the Remainder in terms of differences between terms of "like" kind:

$$\begin{aligned} \text{Rem} &= f(x) - f(x_o) - \nabla f(x_o) \bullet (x - x_o) \\ &= \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(w_k^*)(x_k - x_{ok}) - \nabla f(x_o) \bullet (x - x_o) \\ &= \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(w_k^*)(x_k - x_{ok}) - \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x_o)(x_k - x_{ok}) \\ &= \sum_{k=1}^{n} \left(\frac{\partial f}{\partial x_k}(w_k^*) - \frac{\partial f}{\partial x_k}(x_o) \right)(x_k - x_{ok}). \end{aligned}$$

The last sum is a dot product, so by the Schwarz Inequality

$$\begin{aligned} |\text{Rem}| &= \left| \sum_{k=1}^{n} \left(\frac{\partial f}{\partial x_k}(w_k^*) - \frac{\partial f}{\partial x_k}(x_o) \right)(x_k - x_{ok}) \right| \\ &\le \sqrt{\sum_{k=1}^{n} \left(\frac{\partial f}{\partial x_k}(w_k^*) - \frac{\partial f}{\partial x_k}(x_o) \right)^2} \cdot \sqrt{\sum_{k=1}^{n} (x_k - x_{ok})^2} \\ &= \sqrt{\sum_{k=1}^{n} \left(\frac{\partial f}{\partial x_k}(w_k^*) - \frac{\partial f}{\partial x_k}(x_o) \right)^2} \cdot |x - x_o|. \end{aligned}$$

Thus $|\text{Rem}|/|x - x_o| \le \sqrt{\sum_{k=1}^{n} \left(\frac{\partial f}{\partial x_k}(w_k^*) - \frac{\partial f}{\partial x_k}(x_o) \right)^2} \to 0$ as $x \to x_o$, so f *is* differentiable at x_o.