Higher order partial derivatives

Higher order partial derivatives are defined similarly to higher order derivatives of real functions.

Let $D \subseteq \mathbb{R}^n$ be an open set and $f : D \to \mathbb{R}$ be a real-valued function. Let $i, j = 1, \dots, n$ and $a \in D$. If the i-th partial derivative exists on D, $\frac{\partial f}{\partial x_i} : D \to \mathbb{R}$, we define the <u>ij-th second order partial derivative of f at $a \in D$</u> as the j-th partial derivative of $\frac{\partial f}{\partial x_i}$ at a, provided this exists. Thus, the ij-th partial derivative of f at a is $\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(a)$. This is usually abbreviated $\frac{\partial^2 f}{\partial x_j \partial x_i}(a)$ or $D_j D_i f(a)$.

If $i = j$, then we use the notation $\frac{\partial^2 f}{\partial x_i^2}(a)$.

Let now $k = 1, \dots, n$. Assume that the ij-th second order partial derivative exists on D, $\frac{\partial^2 f}{\partial x_j \partial x_i} : D \to \mathbb{R}$. We define the <u>ijk-th third order partial derivative of f at a</u> as the k-th partial derivative of $\frac{\partial^2 f}{\partial x_j \partial x_i}$ at a, provided that this exists; it is denoted with $\frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i}$ and if it exists for all $a \in D$, we have a function $\frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i} : D \to \mathbb{R}$.

And so on for still higher order partial derivatives.

When still higher order partial derivatives are in question, certain obvious abbreviations are used. For example

$$\frac{\partial^4 f}{\partial x \, \partial y^2 \, \partial z} \quad \text{means} \quad \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial z}\right)\right)\right).$$

f is called twice partial differentiable at a if all second order partial derivatives at a exist. By induction, for $k \in \mathbb{N}$, we say that f is k-times partial differentiable at a if all partial derivatives of order k at a exist.

f is called k-times continuously partial differentiable when f is k-times partial differentiable and all the partial derivatives of order $\leq k$ are continuous.

The large number of possible higher order partial derivatives of a function of several variables is much reduced by the circumstance that the order of performing the partial differentiation is usually irrelevant. The simplest case of this is the equation

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

<u>Theorem 3.54</u> (of H. A. Schwarz)

Let $D \subseteq \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$. Assume that f is twice continuously partial differentiable. Then for all $a \in D$ and for all $i, j = 1, \dots, n$

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a).$$

Proof

Without loss of generality, we can assume that $n=2, i=1, j=2$. Instead of (x_1,x_2) we write (x,y). Thus, we have to prove that for all $a=(x_0,y_0)\in D$,

$$\frac{\partial^2 f}{\partial y\,\partial x}(x_0,y_0)=\frac{\partial^2 f}{\partial x\,\partial y}(x_0,y_0).$$

We consider the sup norm $\|\cdot\|_\infty$ on $\mathbb{R}^2$. Since D is open, there exists $r>0$ s.t. $U_r((x_0,y_0))=\{(x,y)\in\mathbb{R}^2 : |x-x_0|<r$ and $|y-y_0|<r\}\subseteq D$.

For any $(x,y)\in U_r((x_0,y_0))$, consider the expression

$$S(x,y)=f(x,y)-f(x,y_0)-f(x_0,y)+f(x_0,y_0).$$

For $y\in\mathbb{R}$ s.t. $|y-y_0|<r$, define

$$F_y:\,]x_0-r,x_0+r[\,\to\mathbb{R},\quad F_y(x)=f(x,y)-f(x,y_0),$$

so that $S(x,y)=F_y(x)-F_y(x_0)$. By the Mean Value Theorem (Analysis I), there exists η between x_0 and x s.t.

$$F_y(x)-F_y(x_0)=(x-x_0)\cdot F_y'(\eta)=(x-x_0)\left(\frac{\partial f}{\partial x}(\eta,y)-\frac{\partial f}{\partial x}(\eta,y_0)\right)$$

Let us consider now the function

$$G_\eta:\,]y_0-r,y_0+r[\,\to\mathbb{R},\quad G_\eta(y)=\frac{\partial f}{\partial x}(\eta,y)$$

Then we can apply the Mean Value Theorem to this function to get η between y_0 and y s.t.

$$G_\xi(y) - G_\xi(y_0) = (y-y_0)\cdot G_\xi'(\eta) = (y-y_0)\frac{\partial^2 f}{\partial y \partial x}(\xi,\eta).$$

Hence,

$$(1) \qquad S(x,y) = (x-x_0)\cdot(y-y_0)\cdot \frac{\partial^2 f}{\partial y \partial x}(\xi,\eta),$$

where ξ is between x_0 and x, η is between y_0 and y.

A similar argument with the roles of the two variables x and y reversed gives

$$(2) \qquad S(x,y) = (y-y_0)(x-x_0)\frac{\partial^2 f}{\partial x \partial y}(\tilde{\xi},\tilde{\eta}),$$

with $\tilde{\xi}$ between x_0 and x, $\tilde{\eta}$ between y_0 and y.

From (1) and (2) we get for $x \neq x_0$, $y \neq y_0$ that

$$\frac{\partial^2 f}{\partial y \partial x}(\xi,\eta) = \frac{\partial^2 f}{\partial x \partial y}(\tilde{\xi},\tilde{\eta}).$$

Since $\frac{\partial^2 f}{\partial y \partial x}$, $\frac{\partial^2 f}{\partial x \partial y}$ are continuous and $(\xi,\eta)\to(x_0,y_0)$, $(\tilde{\xi},\tilde{\eta})\to(x_0,y_0)$ as $(x,y)\to(x_0,y_0)$, it follows that

$$\frac{\partial^2 f}{\partial y \partial x}(x_0,y_0) = \lim_{(x,y)\to(x_0,y_0)} \frac{\partial^2 f}{\partial y \partial x}(\xi,\eta) =$$

$$= \lim_{(x,y)\to(x_0,y_0)} \frac{\partial^2 f}{\partial x \partial y}(\tilde{\xi},\tilde{\eta}) = \frac{\partial^2 f}{\partial x \partial y}(x_0,y_0). \qquad \square$$

Corollary 9.55

Let $D \subseteq \mathbb{R}^n$ be open and $f: D \to \mathbb{R}$ be k-times continuously partial differentiable. Then

$$\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_2} \partial x_{i_1}} = \frac{\partial^k f}{\partial x_{\pi(k)} \cdots \partial x_{\pi(1)}}$$

for any permutation π of $\{1, \dots, k\}$.

Proof

By induction on k, using Theorem 9.54 $\square$

Definition 9.56

Let $D \subseteq \mathbb{R}^n$ be open and $f: D \to \mathbb{R}$ be twice continuously differentiable. For any $a \in D$, the Hessian matrix of f at a, denoted by $(\mathrm{Hess}\, f)(a)$ is the matrix

$$(\mathrm{Hess}\, f)(a) = \left(\frac{\partial^2 f}{\partial x_j \partial x_i} \right)_{i,j = 1, \dots, n}$$

As a consequence of Theorem 9.54, the Hessian matrix is symmetric.

Example

Let $f: \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1^2 x_2 + x_1 + x_2^3$.

Then f is twice continuously differentiable on D and for every $x = (x_1, x_2) \in D$,

$$\partial f(x) = \left(\frac{\partial f}{\partial x_1}(x) \quad \frac{\partial f}{\partial x_2}(x) \right) = \left(2x_1x_2 + 1 \quad x_1^2 + 3x_2^2 \right),$$

$$\text{grad } f(x) = [\partial f(x)]^T = \begin{pmatrix} 2x_1x_2 + 1 \\ x_1^2 + 3x_2^2 \end{pmatrix}$$

$$\text{Hess } f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) \end{pmatrix}$$

$$= \begin{pmatrix} 2x_2 & 2x_1 \\ 2x_1 & 6x_2 \end{pmatrix}$$

Taylor's Theorem

Our goal in this section is to obtain Taylor approximations for functions of the type $f: D \longrightarrow \mathbb{R}$, where $D \subseteq \mathbb{R}^n$ is an open set.

We shall use the following abbreviations. For a multi-index $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$ and for $x = (x_1, \dots x_n) \in \mathbb{R}^n$, set

$$|\alpha| := \alpha_1 + \dots + \alpha_n \qquad (\text{length of } \alpha)$$

$$\alpha! := \alpha_1! \dots \alpha_n!$$

$$x^\alpha := x_1^{\alpha_1} \dots x_n^{\alpha_n}$$

If f is $|\alpha|$-times continuously partial differentiable, set

$$\partial^\alpha f(x) := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_n^{\alpha_n}}(x) \qquad \text{for every } x \in D,$$

where $\partial x_i^{\alpha_i} = \underbrace{\partial x_i \dots \partial x_i}_{\alpha_i\text{-times}}$.

Lemma 9.57

Let $D \subseteq \mathbb{R}^n$ be open and $f: D \to \mathbb{R}$ be k-times continuously differentiable. Assume that $x \in D$ and $h \in \mathbb{R}^n$ are such that $[x, x+h] \subseteq D$.

Then the function

$$g : [0,1] \to \mathbb{R}, \quad g(t) = f(x+th)$$

is k-times continuously differentiable and

$$g^{(k)}(t) = \sum_{|\alpha|=k} \frac{k!}{\alpha!} \partial^\alpha f(x+th) \cdot h^\alpha$$

Proof

We prove by induction on k that

$$(*) \quad g^{(k)}(t) = \sum_{i_1,\dots,i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i_1}} (x+th) \cdot h_{i_1} \dots h_{i_k}.$$

$k=1$: Use Chain Rule (Corollary 1.36) and the fact that

$g = f \circ \varkappa$, where $\varkappa : [0,1] \to \mathbb{R}^n$, $\varkappa(t) = x + th$,

$\varkappa = \begin{pmatrix} \varkappa_1 \\ \vdots \\ \varkappa_n \end{pmatrix}$, $\varkappa_i(t) = x_i + th_i$, so $\varkappa_i'(t) = h_i$.

It follows that $g'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} (x+th) \cdot h_i$.

$k \Rightarrow k+1$

$$g^{(k+1)}(t) = \frac{\partial}{\partial t} \left(\sum_{i_1,\dots,i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i_1}} (x+th) h_{i_1} \dots h_{i_k} \right)$$

$$= \sum_{j=1}^{n} \frac{\partial}{\partial x_j} \left(\sum_{i_1,\dots,i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i_1}} (x+th) h_{i_1} \dots h_{i_k} \right) h_j$$

$$= \sum_{i_1,\dots,i_k,i_{k+1}=1}^{n} \frac{\partial^{k+1} f}{\partial x_{i_{k+1}} \partial x_{i_k} \dots \partial x_{i_1}}(x+th)\, h_{i_1} \dots h_{i_k} h_{i_{k+1}}.$$

In the following, we shall see how the formula $(*)$ can be simplified using multi-indices.

To any $(i_1,\dots,i_k) \in N^k$ we can associate a multi-index $\alpha = (\alpha_1,\dots,\alpha_n) \in \mathbb{N}_0^n$ by $\alpha_i :=$ the number of appearances of i in $(i_1,\dots,i_k)$. Then $|\alpha| = k$ and, by Corollary 9.55,

$$\frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i}}(x+th)\, h_{i_1} \dots h_{i_k} = \frac{\partial^k f}{\partial x_1^{\alpha_1} \cdot \ldots \cdot \partial x_n^{\alpha_n}} h_1^{\alpha_1} \cdot \ldots \cdot h_n^{\alpha_n}$$

$$= \partial^\alpha f(x+th) \cdot h^\alpha.$$

For any multi-index $\alpha \in \mathbb{N}_0^n$ with $|\alpha| = k$ there are exactly $\frac{k!}{\alpha_1! \cdot \ldots \cdot \alpha_n!}$ k-tuples $(i_1,\dots,i_k)$ such that α is the associated multi-index. It follows that

$$g^{(k)}(t) = \sum_{i_1,\dots,i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i_1}}(x+th)\, h_{i_1} \dots h_{i_k}$$

$$= \sum_{|\alpha|=k} \frac{k!}{\alpha_1! \cdot \ldots \cdot \alpha_n!} \partial^\alpha f(x+th) \cdot h^\alpha$$

$$= \sum_{|\alpha|=k} \frac{k!}{\alpha!} \partial^\alpha f(x+th) \cdot h^\alpha. \qquad \square$$

Theorem 9.58 (Taylor's Theorem)

Let $D \subseteq \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$ be a $(k+1)$-times continuously differentiable function. Assume that $x \in D$ and $h \in \mathbb{R}^n$ are such that $[x, x+h] \subseteq D$.

Then there exists $\theta \in [0,1]$ such that

$$f(x+h) = \underbrace{\sum_{|\alpha| \le k} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha}_{\text{the Taylor polynomial of order } k \text{ of } f \text{ at } x} + \sum_{|\alpha| = k+1} \frac{\partial^\alpha f(x+\theta h)}{\alpha!} \cdot h^\alpha .$$

Proof

Let us consider the function $g : [0,1] \to \mathbb{R}$, $g(t) = f(x+th)$.

By Lemma 9.57, g is $(k+1)$-times continuously differentiable and

$$g^{(m)}(t) = \sum_{|\alpha| = m} \frac{m!}{\alpha!} \partial^\alpha f(x+th) \cdot h^\alpha . \quad \text{for all } m = 1, \dots, k+1.$$

Applying Taylor's Theorem from Analysis I to g we get that there exists $\theta \in [0,1]$ s.t.

$$g(1) = \sum_{m=0}^{k} \frac{g^{(m)}(0)}{m!} + \frac{g^{(k+1)}(\theta)}{(k+1)!}$$

$$= \sum_{m=0}^{k} \sum_{|\alpha| = m} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha + \sum_{|\alpha| = k+1} \frac{\partial^\alpha f(x+\theta h)}{\alpha!} \cdot h^\alpha$$

$$= \sum_{|\alpha| \le k} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha + \sum_{|\alpha| = k+1} \frac{\partial^\alpha f(x+\theta h)}{\alpha!} \cdot h^\alpha .$$

$\square$

Corollary 9.53

Let $D \subseteq \mathbb{R}^n$ be open, $x \in D$ and $\delta > 0$ be such that $U_\delta(x) \subseteq D$. Assume that $f: D \to \mathbb{R}$ is k-times continuously differentiable. Then for all $h \in \mathbb{R}^n$ with $\|h\| < \delta$,

$$f(x+h) = \sum_{|\alpha| \leq k} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha + \varphi(h),$$

where $\varphi: U_\delta(0) \to \mathbb{R}$ is such that $\varphi(0) = 0$ and $\lim\limits_{\substack{h \to 0 \\ h \neq 0}} \frac{\varphi(h)}{\|h\|^k} = 0$.

Proof

Let $h \in \mathbb{R}^n$ with $\|h\| < 1$. Then $[x, x+h] \subseteq U_\delta(x) \subseteq D$, so we can apply Taylor's Theorem to get $\theta \in [0,1]$ s.t.

$$f(x+h) = \sum_{|\alpha| \leq k-1} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \sum_{|\alpha| = k} \frac{\partial^\alpha f(x+\theta h)}{\alpha!} h^\alpha$$

$$= \sum_{|\alpha| = k} \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha + \varphi_\alpha(h),$$

where $\varphi_\alpha(h) := \frac{\partial^\alpha f(x+\theta h) - \partial^\alpha f(x)}{\alpha!}$.

Since $\partial^\alpha f$ is continuous, we get that $\lim\limits_{h \to 0} \partial^\alpha f(x+\theta h) = \partial^\alpha f(x)$,

so $\lim\limits_{h \to 0} \frac{\partial^\alpha f(x+\theta h) - \partial^\alpha f(x)}{\alpha!} = 0$, that is $\lim\limits_{h \to 0} \varphi_\alpha(h) = 0$.

Let us define $\varphi: U_\delta(0) \to \mathbb{R}$, $\varphi(h) = \sum\limits_{|\alpha| = k} \varphi_\alpha(h) \cdot h^\alpha$.

Then $f(x+h) = \sum_{|\alpha|\leq k} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha + \varphi(h)$, $\varphi(0)=0$

and

$$\frac{|\varphi(h)|}{\|h\|^k} = \frac{\left|\sum_{|\alpha|=k} \varphi_\alpha(h)\cdot h^\alpha\right|}{\|h\|^k} \leq \sum_{|\alpha|=k} |\varphi_\alpha(h)| \cdot \frac{|h^\alpha|}{\|h\|^k} =$$

$$= \sum_{|\alpha|=k} |\varphi_\alpha(h)| \cdot \frac{|h_1^{\alpha_1}| \cdot \ldots \cdot |h_n^{\alpha_n}|}{\|h\|^{\alpha_1}\cdot \ldots \cdot \|h\|^{\alpha_n}} \leq \sum_{|\alpha|=k} |\varphi_\alpha(h)|,$$

since $|h_i| \leq \|h\|$ for all $i=1,\ldots,n$.

Since $\lim_{h\to 0} \varphi_\alpha(h) = 0$, it follows that $\lim_{h\to 0} \frac{\varphi(h)}{\|h\|^k} = 0$. □

Let us recall that the canonical scalar product on $\mathbb{R}^n$ is defined by:

$$\langle x, y\rangle = \sum_{i=1}^n x_i y_i \qquad \text{for all } x = \begin{pmatrix} x_1 \\ \vdots \\ x_n\end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_n\end{pmatrix} \in \mathbb{R}^n.$$

<u>Corollary 3.60</u>

Let $D \subseteq \mathbb{R}^n$ be open, $x\in D$ and $\delta>0$ be such that $U_\delta(x)\subseteq D$.

Assume that $f: D \to \mathbb{R}$ is continuously differentiable.

Then for all $h\in\mathbb{R}^n$ with $\|h\|<\delta$,

$$f(x+h) = f(x) + \langle \operatorname{grad} f(x), h\rangle + \varphi(h),$$

where $\varphi: U_\delta(0) \to \mathbb{R}$ is s.t. $\varphi(0)=0$ and $\lim_{h\to 0} \frac{\varphi(h)}{\|h\|} = 0$.

Proof.

Apply Corollary 9.59 with $k=1$. We get that

$$f(x+h) = \sum_{|\alpha|\leq 1} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha + \varphi(h),$$

where $\varphi: U_\delta(0) \to \mathbb{R}$, $\varphi(0)=0$ and $\lim_{h\to 0} \frac{\varphi(h)}{\|h\|} = 0$.

We have the following cases:

1) $|\alpha| = 0$. The only multi-index $\alpha \in \mathbb{N}_0^n$ with $|\alpha|=0$ is $\alpha = (0,\dots,0)$. It follows that

$$\sum_{|\alpha|=0} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha = \frac{\partial^0 f(x)}{0!} \cdot h^0 = f(x).$$

2) $|\alpha| = 1$. The multi-indices $\alpha \in \mathbb{N}_0^n$ with $|\alpha|=1$ are the unit vectors

$$e_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \to j\text{-th position}$$

It follows that

$$\sum_{|\alpha|=1} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha = \sum_{j=1}^n \frac{\partial^{e_j} f(x)}{e_j!} \cdot h^{e_j} =$$

$$= \sum_{j=1}^n \frac{\partial f}{\partial x_j}(x) \cdot h_j = \langle \text{grad}\, f(x), h \rangle, \text{ since}$$

$$\partial^{e_j} f(x) = \frac{\partial f}{\partial x_j}(x), \quad e_j! = 1 \text{ and } h^{e_j} = h_j.$$

Thus, $f(x+h) = f(x) + \langle \text{grad}\, f(x), h\rangle + \varphi(h)$.

$\square$

Corollary 9.61

Let $D \subseteq \mathbb{R}^n$ be open, $x \in D$ and $\delta > 0$ be such that $U_\delta(x) \subseteq D$.
Assume that f is twice continuously differentiable.
Then for all $h \in \mathbb{R}^n$ with $\|h\| < \delta$,

$$f(x+h) = f(x) + \langle \operatorname{grad} f(x), h \rangle + \frac{1}{2} \langle h, (\operatorname{Hess} f)(x) \cdot h \rangle + \varphi(h),$$

where $\varphi : U_\delta(0) \to \mathbb{R}$ is such that $\varphi(0) = 0$ and $\lim_{h \to 0} \frac{\varphi(h)}{\|h\|^2} = 0$.

Proof

Apply Corollary 9.59 with $k = 2$. We get that

$$f(x+h) = \sum_{|\alpha| \leq 1} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha + \varphi(h),$$

where $\varphi : U_\delta(0) \to \mathbb{R}$, $\varphi(0) = 0$ and $\lim_{h \to 0} \frac{\varphi(h)}{\|h\|^2} = 0$.

By the proof of Corollary 9.60, it follows that

$$f(x+h) = f(x) + \langle \operatorname{grad} f(x), h \rangle + \sum_{|\alpha| = 2} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha + \varphi(h).$$

We have to find out all the multi-indices $\alpha \in \mathbb{N}_0^n$ with $|\alpha| = 2$. There are two types of such multi-indices:

1) $2e_i = (0, \ldots 0, 2, 0, \ldots 0)$ $\quad 1 \leq i \leq n$.

$$\partial^{2e_i} f(x) = \frac{\partial^2 f}{\partial x_i^2}(x), \quad (2e_i)! = 2, \quad h^{2e_i} = h_i^2$$

2) $e_i + e_j = (0, \ldots, 1, \ldots, 1, \ldots 0)$ $\quad 1 \leq i < j \leq n$.

$$\partial^{e_i+e_j} f(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x), \quad (e_i+e_j)! = 1, \quad h^{(e_i+e_j)} = h_i h_j.$$

Hence,

$$\sum_{|\alpha|=2} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha = \frac{1}{2}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}(x) \cdot h_i^2 + \\ + \sum_{1\le i<j\le n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \cdot h_i \cdot h_j$$

Since, by Schwarz' Theorem 8.54, we have that

$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$, it follows that

$$\sum_{1\le i<j\le n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \cdot h_i h_j = \frac{1}{2} \sum_{1\le i\ne j\le n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) h_i h_j.$$

Then

$$\sum_{|\alpha|=2} \frac{\partial^\alpha f(x)}{\alpha!} \cdot h^\alpha = \frac{1}{2}\left(\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}(x) \cdot h_i^2 + \sum_{i\ne j} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \cdot h_i h_j \right)$$

$$= \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) h_i h_j$$

$= \frac{1}{2} \langle h, Ah \rangle$, where $A = \text{Hess} f(x)$. $\square$