1 Taylor series with remainder

$$\phi(t), \phi^{(1)}(t), \phi^{(n)}(t), \phi_i \in \mathbb{R} \to \mathbb{R}$$

exist and are continous for $\forall t \in (a,b), [x, x+h] \subset (a,b)$
then:

$$\phi(x+h) = \phi(x) + \phi^{(1)}(x)h + .. + \phi^{(n)}(x)\frac{h^n}{n!} + ..$$

for some $\hat{x} \in (x, x+h)$

$$f : \mathbb{R}^n \to \mathbb{R}, \nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ .. \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ ..x_n \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & .. \\ .. & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{bmatrix}$$

$f^{(2)}(x)$ is continuous $\implies$ mixed partial derivatives are commutative
$f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable $\implies$
$f(x+p) = f(x) + \nabla f(x+tp)^T p, \exists t \in (0,1)$

$$\begin{aligned}
&\text{let } \phi(t) = f(x+tp) \\
\phi'(t) &= \lim_{h\to 0} \frac{\phi(t+h) - \phi(t)}{h} = \lim_{h\to 0} \frac{f(x+(t+h)p) - f(x+tp)}{h} \\
&\text{let } \ x \in \mathbb{R}^2 \\
\phi'(t) &= \lim_{h\to 0} \frac{f\left(\begin{bmatrix} x_1 + (t+h)p_1 \\ x_2 + (t+h)p_2 \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1 + tp_1 \\ x_2 + tp_2 \end{bmatrix}\right)}{h} \\
&= \lim_{h\to 0} \frac{f\left(\begin{bmatrix} x_1 + (t+h)p_1 \\ x_2 + (t+h)p_2 \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1 + tp_1 \\ x_2 + (t+h)p_2 \end{bmatrix}\right)}{h} \\
&= \lim_{h\to 0} \frac{f\left(\begin{bmatrix} x_1 + (t+h)p_1 \\ x_2 + (t+h)p_2 \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1 + tp_1 \\ x_2 + (t+h)p_2 \end{bmatrix}\right)}{h} \\
&+ \lim_{h\to 0} \frac{f\left(\begin{bmatrix} x_1 + tp_1 \\ x_2 + (t+h)p_2 \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1 + tp_1 \\ x_2 + tp_2 \end{bmatrix}\right)}{h}
\end{aligned}$$

$$
\begin{aligned}
\phi'(t) &= \lim_{h\to 0} \frac{f\left(\begin{bmatrix} x_1+(t+h)p_1 \\ x_2+(t+h)p_2 \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1+tp_1 \\ x_2+(t+h)p_2 \end{bmatrix}\right)}{hp_1}\frac{hp_1}{h} \\
&+ \lim_{h\to 0} \frac{f\left(\begin{bmatrix} x_1+tp_1 \\ x_2+(t+h)p_2 \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1+tp_1 \\ x_2+tp_2 \end{bmatrix}\right)}{hp_2}\frac{hp_2}{h} \\
&= \frac{\partial f}{\partial x_1}(x+tp)p_1 + \frac{\partial f}{\partial x_2}(x+tp)p_2 \\
f(x+tp) &= f(x) + \nabla f(x+tp)^T p \\
\phi(1) &= \phi(0) + \phi'(t) \\
\phi''(t) &= p^T \nabla^2 f(x+tp)p \\
\phi(1) &= \phi(0) + \phi'(0) + \frac{\phi''(t)}{2}, \text{ for some } t \in (0,1) \\
f(x+p) &= f(x) + \nabla f(x)^T p + \frac{1}{2}p^T \nabla f(x+tp)p, f : \mathbb{R}^n \to \mathbb{R} \\
\nabla f(x+p) &= \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)pdt \\
\phi(t) &= \nabla f(x+tp)_i \in \mathbb{R} \\
\phi(1) &= \phi(0) + \int_o^1 \phi'(t)dt \\
\nabla f(x+tp)_i &= \nabla f(x)_i + \int_0^1 \nabla^2 f(x+tp)_i^T pdt, \text{ i=ith row of Hessian matrix} \\
\nabla f(x+tp) &= \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)^T pdt
\end{aligned}
$$

note:
$\nabla f(x+p)_i = \nabla f(x)_i + \nabla^2 f(x+t_i p)p$, for some $t_i \in (0,1)$ is true
$\nabla f(x+p) = \nabla f(x) + \nabla^2 f(x+tp)p$, for some $t \in (0,1)$ is not true

2 1st order conditions

Theorem 2.1. *1st order necessary condition.*
if x^* is a local minimizer of f and f is continuously differentiable in a beighbourhood of x^* then $\nabla f(x^*) = 0$

Proof. By contradiction (assume hypothesis true, conclusion false)
let $p = -\nabla f(x^*)$
note $\nabla f(x^*)^T p = -\nabla f(x^*)^T \nabla f(x^*) < 0$
consider $-\nabla f(x^*)^T \nabla f(x * + tp) < 0$ for $t \in [0, \alpha], \alpha > 0$
$\phi(t) = \nabla f(x^*)^T \nabla f(x^* + tp), \phi : \mathbb{R} \to \mathbb{R}$
$\phi(0) = \nabla f(x^*)^T \nabla f(x^*) > 0$ (using assumed false conclusion)
$|\phi(o) - \phi(t)| < \epsilon = \frac{\phi(0)}{2}, |t| < \delta$
by continuity of $\phi(t)$,
$(\exists \alpha > 0) - \nabla f(x^*)^T \nabla f(x^* + tp) < 0, t \in [0, \alpha]$
$f(x^* + \alpha p) = f(x^*) + \alpha \nabla f(x + tp)^T p$, for some $t \in [0, \alpha]$
$\alpha \nabla f(x + tp)^T p < 0$, then:
$f(x^* + \alpha p) < f(x^*)$
x^* is a local minimizer
there esits an $r > 0$ s.t. for all $x(x^*, r) f(x^*)(x)$
note for $x = x^* + \alpha p : \|x - x^*\| = |a| \|p\| < r$
we can adjust t to all smaller values and $\alpha \nabla f(x + tp)^T p < 0$ would still hold
thus contradiction □