Problem 1:

If $f(x_1, x_2) = x_1^3 + x_2^3$

then $\nabla f = (3x_1^2, 3x_2^2)$

Let $\nabla f = 0 \Rightarrow (x_1, x_2) = (0, 0)$

And for the second derivative:

$$\nabla^2 f = \begin{pmatrix} 6x_1 & 0 \\ 0 & 6x_2 \end{pmatrix}$$

When $x^* = (x_1, x_2) = (0, 0)$

we will have

$$\nabla^2 f(x^*) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$\nabla^2 f(x^*)$ is a positive semi-definite matrix because:

$$\forall x \neq 0, \quad x^T \nabla^2 f(x^*) x = 0 \geq 0$$

Then we could tell that

$x^* = (0, 0)$ satisfies the second-order necessary conditions for optimality

However, as for points near x^* we could choose:

$$x_1 = (\varepsilon, \varepsilon) \text{ and}$$

$$x_2 = (-\varepsilon, -\varepsilon), \ \forall \varepsilon > 0$$

Then $f(x_1) = 2\varepsilon^3 > 0$

$f(x_2) = -2\varepsilon^3 < 0$

So x^* is not a local optimum

As a result,

$$f(x_1, x_2) = x_1^3 + x_2^3$$

is an example!

Problem 2:

$\forall x_1, x_2 \in L(y) = \{x \in \mathbb{R}^n : f(x) \leq y\}$

$f(x_1) \leq y$ and $f(x_2) \leq y$

Since $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ is a convex function,

then $\forall \lambda \in [0, 1]$

we have

$$
\begin{aligned}
f[(1-\lambda)x_1 + \lambda x_2] &\leq (1-\lambda) f(x_1) + \lambda f(x_2) \\
&\leq (1-\lambda) y + \lambda y \\
&= y
\end{aligned}
$$

So we prove that

$$(1-\lambda)x_1 + \lambda x_2 \in L(y)$$

This means

$$L(y) = \{x \in \mathbb{R}^n : f(x) \leq y\}$$

is a convex set.

Problem 3 :

$\forall x' \in L(y)$

we have

$$f(x') \le y = f(x)$$

Since $f: \mathbb{R}^n \to \mathbb{R}$ is convex

then

$$f(x') \ge f(x) + \nabla f(x)^T (x' - x), \quad (*)$$

From $(*)$ we know that

$$f(x) \ge f(x') \ge f(x) + \nabla f(x)^T (x' - x)$$

$$\Rightarrow \nabla f(x)^T (x' - x) \le 0$$

$$\Rightarrow x' \in H(x)$$

That is to say,

$\forall\, x' \in L(y)$,

we always have $x' \in H(x)$

As a result,

$$L(y) \subseteq H(x)$$

By the way, we could prove the equation $(*)$ as below:

let $z = t x' + (1-t) x$

then $z = x + t(x' - x)$

Since f is convex

$$\begin{aligned} \Rightarrow f(z) &= f[x + t(x' - x)] \\ &= f[t x' + (1-t) x] \\ &\le t f(x') + (1-t) f(x) \\ &= f(x) + t[f(x') - f(x)] \end{aligned}$$

That is to say,

$$f[x+t(x'-x)] \le f(x) + t[f(x')-f(x)]$$

$$\Leftrightarrow \frac{f[x+t(x'-x)]-f(x)}{t} \le f(x')-f(x)$$

$$\Leftrightarrow \frac{f[x+t(x'-x)]-f(x)}{t(x'-x)}(x'-x) \le f(x')-f(x)$$

Let $t \to 0$ we will get

$$\nabla f(x)^T (x'-x) \le f(x')-f(x)$$

Therefore

$$f(x') \ge f(x) + \nabla f(x)^T (x'-x)$$

Problem 4:

Since $f: \mathbb{R}^n \to \mathbb{R}$ is a strictly convex function,

then $\nabla^2 f(x)$ is positive definite, $\forall x \in \mathbb{R}^n$

$\forall x' \in L(f(x))$, we have

$$f(x') \leq f(x)$$

And we could do Taylor expansion to f:

$$f(x') = f(x) + \nabla f(x)^T (x' - x) + \frac{1}{2}(x'-x)^T \nabla^2 f(\xi)(x'-x)$$

if $x' \neq x$ then

$$(x'-x)^T \nabla^2 f(\xi)(x'-x) > 0$$

That is to say,

$$f(x') = f(x) + \nabla f(x)^T (x'-x) + \frac{1}{2}(x'-x)^T \nabla^2 f(a)(x'-x)$$

$$> f(x) + \nabla f(x)^T (x'-x)$$

Then we have:

$$f(x) \geq f(x') > f(x) + \nabla f(x)^T (x'-x)$$

$$\Downarrow$$

$$\nabla f(x)^T (x'-x) < 0$$

But if $x' = x$

then $\nabla f(x)^T (x'-x) = 0$

$$\Downarrow$$

$$L(f(x)) \cap \{x \in \mathbb{R}^n : \nabla f(x)^T (x'-x) = 0\} = \{x\}$$

Problem 5:

Since $p \in \mathbb{R}^n$ is a descent direction at x, then we have:

$$\nabla f(x)^T p < 0$$

Let's do Taylor expansion to f:

$$f(x+\alpha p) = f(x) + \alpha \nabla f(x)^T p + \frac{\alpha^2}{2} p^T \nabla^2 f(\xi) p$$

if $f(x+\alpha p) < f(x)$

then $f(x) + \alpha \nabla f(x)^T p + \frac{\alpha^2}{2} p^T \nabla^2 f(\xi) p < f(x)$

$$\Downarrow$$

$$\alpha \nabla f(x)^T p + \frac{\alpha^2}{2} p^T \nabla^2 f(\xi) p < 0$$

$$\Rightarrow \frac{\alpha^2}{2} p^T \nabla^2 f(\xi) p < \alpha(-\nabla f(x)^T p)$$

$$\Rightarrow \alpha < \frac{-2\nabla f(x)^T p}{p^T \nabla^2 f(\xi) p}$$

since $\nabla f(x)^T p < 0$ and $\nabla^2 f(\xi)$ is positive definite

$$\Downarrow$$

$$\frac{-2\nabla f(x)^T p}{p^T \nabla^2 f(\xi) p} > 0$$

That is to say, $\alpha \in \left(0, \frac{-2\nabla f(x)^T p}{p^T \nabla^2 f(\xi) p}\right)$

Therefore, there exists $\alpha > 0$ such that $f(x+\alpha p) < f(x)$

Problem 6:

① We know that $q_1, \cdots, q_n$ form an orthonormal basis for $\mathbb{R}^n$

Then $q_i^T q_j = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$

Let $Q = [q_1 \ q_2 \ \cdots \ q_n]$

we have $QQ^T = I$

$$\Downarrow$$

$$Q^{-1} = Q^T$$

Meanwhile,

since we have $Aq_i = \lambda_i q_i$, $i = 1, 2, \cdots, n$

then $AQ = [Aq_1 \ Aq_2 \ \cdots \ Aq_n]$

$= [\lambda_1 q_1 \ \lambda_2 q_2 \ \cdots \ \lambda_n q_n]$

$$= [q_1 \; q_2 \cdots q_n] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

$$= Q\Lambda$$

That is to say,

$$AQ = Q\Lambda$$

$$\Downarrow$$

$$A = Q\Lambda Q^{-1} = Q\Lambda Q^T$$

Therefore we prove that

a consequence of this is having an orthogonal eigenvalue decomposition of A:

$$A = Q\Lambda Q^T, \; Q = [q_1 \cdots q_n], \; \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

where $Q^{-1} = Q^T$

② $A = Q \Lambda Q^T$

$$= \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 q_1 & \cdots & \lambda_n q_n \end{bmatrix} \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix}$$

$$= \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots \lambda_n q_n q_n^T$$

$$= \sum_{i=1}^{n} \lambda_i q_i q_i^T$$

So we show that we could write the eigenvalue decomposition in outer product form, i.e.:

$$A = \sum_{i=1}^{n} \lambda_i q_i q_i^T$$

③ (i) if $\lambda_i > 0$ $(i=1,2,\cdots,n)$

then $\forall x \neq 0$ we have

$$x^T A x = x^T \left(\sum_{i=1}^{n} \lambda_i q_i q_i^T\right) x$$

$$= \sum_{i=1}^{n} \lambda_i (x^T q_i)(q_i^T x)$$

$$= \sum_{i=1}^{n} \lambda_i (q_i^T x)^T (q_i^T x)$$

$$= \sum_{i=1}^{n} \lambda_i \| q_i^T x \|^2$$

since $q_1 \cdots q_n$ form an orthonormal basis for $\mathbb{R}^n$

then there is no possibility that

$q_i^T x = 0$, $\forall i \in \{1,2,\cdots,n\}$

This means

there always exists $j \in \{1,2,\cdots,n\}$

s.t. $q_j^T x \neq 0$

That is to say

$$\sum_{i=1}^{n} \lambda_i \| q_i^T x \|^2 \geq \lambda_j \| q_j^T x \|^2 > 0$$

$$\Downarrow$$

$$x^T A x > 0 , \quad \forall x \neq 0$$

so A is positive definite

(iv) if A is positive definite

then $\forall x \neq 0$, $x^T A x > 0$

let's choose $x = q_k$, $k = 1, 2, \cdots, n$

then

$$x^T A x = x^T \left(\sum_{i=1}^{n} \lambda_i q_i q_i^T \right) x$$

$$= q_k^T \left(\sum_{i=1}^{n} \lambda_i q_i q_i^T \right) q_k$$

$$= \sum_{i=1}^{n} \lambda_i (q_i^T q_k)^T (q_i^T q_k)$$

$$= \sum_{i=1}^{n} \lambda_i \| q_i^T q_k \|^2$$

And we know the truth:

$$q_i^T q_k = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases}$$

Then $x^T A x = \sum_{i=1}^{n} \lambda_i \| q_i^T q_k \|^2$

$= \lambda_k > 0, \quad k = 1, 2, \cdots, n$

From (i) and (ii),

we prove that

$\lambda_i > 0 \ (i = 1, 2, \cdots, n)$ if and only if

A is positive definite.

Problem 7:

$\forall x' \in S$ and $x' \neq x^*$ ($x^* \in S$)

Since S is a convex set

then $x^* + \xi(x' - x^*) = \xi x' + (1-\xi)x^* \in S$

(This is the property of convex set!)

$$f(x') = f(x^*) + \nabla f(x^*)^T (x' - x^*)$$

$$+ \frac{1}{2}(x' - x^*)^T \nabla^2 f[x^* + \xi(x' - x^*)](x' - x^*)$$

Since $x^* \in S$ is a local minimum

then $\nabla f(x^*) = 0$

$\Downarrow$

$$f(x') = f(x^*) + \frac{1}{2}(x' - x^*)^T \nabla^2 f[x^* + \xi(x' - x^*)](x' - x^*)$$

Now we have

$\nabla^2 f(x)$ is positive definite

for all $x \in S$

then

$$\frac{1}{2}(x'-x^*)^T \nabla^2 f\left[x^* + \xi(x'-x^*)\right](x'-x^*) > 0$$

$$\Downarrow$$

$$f(x') > f(x^*), \quad \forall x' \in S \text{ and } x' \neq x^*$$

Therefore,

x^* is a global minimum of f over S.

Problem 8

Let $y = Ax + b$

then $g(x) = f(Ax+b)$

So applying Newton step on g will be

$$x' = x - (\nabla^2 g(x))^{-1} \nabla g(x)$$

$$= x - (A^T \nabla^2 f(Ax+b) A)^{-1} A^T \nabla f(Ax+b)$$

$$= x - A^{-1} (\nabla^2 f(Ax+b))^{-1} \nabla f(Ax+b)$$

Then

$$Ax' + b = (Ax+b) - (\nabla^2 f(Ax+b))^{-1} \nabla f(Ax+b)$$

$$\Updownarrow$$

$$y' = y - (\nabla^2 f(y))^{-1} \nabla f(y)$$

That is to say
Newton's method has the property of affine invariance

Therefore,
the Newton step for $g(x)=f(Ax+b)$ is the same as the Newton step for $f(x)$