6.6 Let $f(x,y) = 3x^2y + 4xy^2 + xy$. Observe that

$$Df(x,y) = [6xy + 4y^2 + y, 3x^2 + 8xy + x]$$

Now we find the critical points.
Case 1. $x = 0$:

$$\begin{aligned} 0 &= 4y^2 + y \\ &= y(4y+1) \end{aligned}$$

and $y = 0$ or $y = -\frac{1}{4}$

Case 2. $y = 0$:

$$\begin{aligned} 0 &= 3x^2 + x \\ &= x(3x+1) \end{aligned}$$

and $x = 0$ or $x = -\frac{1}{3}$

Case 3. $(x,y) \neq (0,0)$:

$$\begin{aligned} 0 &= 6xy + 4y^2 + y \\ &\Leftrightarrow \\ x &= \frac{(-4y^2 - y)}{6y} \\ &= \frac{-4y-1}{6} \end{aligned}$$

So, observe that

$$\begin{aligned}
0 &= 3x^2 + 8xy + x \\
&= 3\left(\frac{-4y-1}{6}\right)^2 + 8\left(\frac{-4y-1}{6}\right)y + \left(\frac{-4y-1}{6}\right) \\
&= \left(\frac{-4y-1}{6}\right)\left(3\frac{(-4y-1)}{6} + 8y + 1\right)
\end{aligned}$$

Now if $\frac{-4y-1}{6} = 0$, then $y = -\frac{1}{4}$ which implies that $x = 0$ by *Case 1*. However, we assumed in this case that $x \neq 0$. So we check where $\left(3\frac{(-4y-1)}{6}+8y+1\right) = 0$. Observe that

$$\begin{aligned}
0 &= \left(3\frac{(-4y-1)}{6} + 8y + 1\right) \\
&= \frac{-12y - 3 + 48y + 6}{6} \\
&= \left(\frac{36y+3}{6}\right)
\end{aligned}$$

and $y = -\frac{1}{12}$ So

$$\begin{aligned}
x &= \frac{-4y-1}{6} \\
&= \frac{-4(-\frac{1}{12}) - 1}{6} \\
&= \frac{\frac{1}{3} - 1}{6} \\
&= \frac{-\frac{2}{3}}{6} \\
&= -\frac{1}{9}
\end{aligned}$$

Thus, the critical points are: $(0,0)$, $(0,-\frac{1}{4})$, $(-\frac{1}{3},0)$, $(-\frac{1}{9},-\frac{1}{12})$.
We now determine which of these points are the locations of local maxima, minima, or saddle points. Recall that $f(x,y) = 3x^2y + 4xy^2 - xy$. Observe

that

$$
\begin{aligned}
f_x(x,y) &= 6xy + 4y^2 + y \\
f_{xx}(x,y) &= 6y \\
f_y(x,y) &= 3x^2 + 8xy + x \\
f_{yy}(x,y) &= 8x \\
f_{xy}(x,y) &= f_{yx}(x,y) = 6x + 8y + 1 \quad \text{(by Clairut's Theorem)}
\end{aligned}
$$

Now we check each critical point. We will check as follows: first we will compute $\det D^2 f = \det \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix}$ at each critical point. If this is negative, the critical point is a saddle point. If it is positive, and $f_{xx} < 0$, it is a local maxima. If it is positive and $f_{xx} > 0$ it is a local minima. Now,

$(0,0)$: $f_{xx} = 0$, $f_{yy} = 0$, $f_{xy} = 1$. Hence $\det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -1 < 0 \Rightarrow$ $\boxed{\text{saddle point}}$

$(0, -\frac{1}{4})$: $\det \begin{bmatrix} -\frac{3}{2} & -1 \\ -1 & 0 \end{bmatrix} = -1 < 0 \Rightarrow \boxed{\text{saddle point}}$

$(-\frac{1}{3}, 0)$: $\det \begin{bmatrix} 0 & -1 \\ -1 & -\frac{8}{3} \end{bmatrix} = -1 < 0 \ \boxed{\text{saddle point}}$

$(-\frac{1}{9}, -\frac{1}{12})$: $\det \begin{bmatrix} -\frac{1}{2} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{8}{9} \end{bmatrix} = \frac{1}{3} > 0$ and $f_{xx} = -\frac{1}{2} < 0 \Rightarrow \boxed{\text{local maxima}}$

6.7

(i) *Proof.* Let A be a square matrix, and $Q = A^T + A$. Observe that

$$
\begin{aligned}
Q^T &= (A^T + A)^T \\
&= A + A^T \\
&= A^T + A \\
&= Q
\end{aligned}
$$

Since $Q = Q^T$, Q is symmetric. Observe that

$$\begin{aligned}
\mathbf{x}^T Q\mathbf{x} &= \mathbf{x}^T(A^T + A)\mathbf{x} \\
&= (\mathbf{x}^T A^T + \mathbf{x}^T A)\mathbf{x} \\
&= \mathbf{x}^T A^T\mathbf{x} + \mathbf{x}^T A\mathbf{x} \\
&= \mathbf{x}^T A^T\mathbf{x} + (\mathbf{x}^T A^T\mathbf{x})^T
\end{aligned}$$

where $(\mathbf{x}^T A^T\mathbf{x})$ is the same scalar as $(\mathbf{x}^T A^T\mathbf{x})^T$. Hence $\mathbf{x}^T A^T\mathbf{x} + \mathbf{x}^T A\mathbf{x} = 2\mathbf{x}^T A\mathbf{x}$. So

$$\begin{aligned}
\mathbf{x}^T Q\mathbf{x} &= 2\mathbf{x}^T A\mathbf{x} \\
\frac{1}{2}\mathbf{x}^T Q\mathbf{x} &= \mathbf{x}^T A\mathbf{x} \\
\frac{1}{2}Q &= A
\end{aligned}$$

Thus

$$\begin{aligned}
f(x) &= \mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x} + c \\
&= \mathbf{x}^T\frac{1}{2}Q\mathbf{x} - \mathbf{b}^T\mathbf{x} + c \\
&= \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{b}^T\mathbf{x} + c
\end{aligned}$$

$\square$

(ii) Let $\mathbf{x}^*$ be a minimizer of f. Observe that

$$\begin{aligned}
f(x) &= \mathbf{x}^T A\mathbf{x} - \mathbf{b}^T + c \\
Df(x) = \mathbf{x}^{*T}(A + A^T) - \mathbf{b}^T &= 0 \\
\mathbf{x}^{*T}(A + A^T) &= \mathbf{b}^T \\
(\mathbf{x}^{*T}(A + A^T))^T &= \mathbf{b} \\
(A + A^T)^T\mathbf{x}^* &= \mathbf{b} \\
Q^T\mathbf{x}^* &= \mathbf{b} \quad (\text{since } Q = Q^T)
\end{aligned}$$

(iii)

6.11

Proof. Let $f(x) = ax^2 + bx + c$, where $a > 0$ and $b, c \in \mathbb{R}$. Let $x_0 \in \mathbb{R}$ be an initial guess. Observe that

$$f(x) = a\left(x + \frac{b}{2a}\right)^2 - a\left(\frac{b}{2a}\right)^2 + c$$

hence the minimum is achieved at $x = -\frac{b}{2a}$.Note that one iteration of Newton's method yields the following:

$$\begin{aligned} x_1 &= x_0 - \frac{f'(x_0)}{f''(x_0)} \\ &= x_0 - \frac{2ax_0 + b}{2a} \\ &= -\frac{b}{2a} \end{aligned}$$

Therefore, one iteration of Newton's method lands at the unique minimizer of f. $\square$

6.15 See Jupyter notebook.
7.1

Proof. Let S be a nonempty subset of V. By definition, the *convex hull* of S, denoted $\text{conv}(S)$ is the set of all convex combinations of elements of S; that is, the set of all finite sums of the form

$$\lambda_1 \mathbf{x}_1 + ... + \lambda_k \mathbf{x}_k, \quad \mathbf{x}_i \in S, k \in \mathbb{N}$$

where each $\lambda_i \geq 0$ and $\lambda_1 + ... + \lambda_k = 1$ $\square$

7.2

(i) *Proof.* Suppose we have a hyperplane in V, which by definition is a set of the form $P = \{\mathbf{x} \in V \mid \langle \mathbf{a}, \mathbf{x} \rangle\} = b$ where $\mathbf{a} \in V$, $\mathbf{a} \neq \mathbf{0}$, and $b \in \mathbb{R}$.

So, suppose $\mathbf{x}, \mathbf{y} \in P$. Observe that

$$\begin{aligned} \langle \mathbf{a}, \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \rangle &= \lambda\langle \mathbf{a}, \mathbf{x} \rangle + (1-\lambda)\langle \mathbf{a}, \mathbf{y} \rangle \\ &= \lambda b + (1-\lambda)b = b \end{aligned}$$

Thus $\lambda\mathbf{x} + (1-\lambda)\mathbf{x} \in P$. Therefore a hyperplane is convex. □

(ii) *Proof.* Suppose we have a half space, which by definition is a set of the form $H = \{\mathbf{x} \in V \mid \langle \mathbf{a}, \mathbf{x} \rangle \leq b\}$, where $\mathbf{a} \in V$, $\mathbf{a} \neq \mathbf{0}$, and $b \in \mathbb{R}$. Suppose that $\mathbf{x}, \mathbf{y} \in H$. Observe that

$$\begin{aligned} \langle \mathbf{a}, \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \rangle &= \lambda\langle \mathbf{a}, \mathbf{x} \rangle + (1-\lambda)\langle \mathbf{a}, \mathbf{y} \rangle \\ &\leq \lambda b + (1-\lambda)b = b \end{aligned}$$

Thus $\lambda\mathbf{x} + (1-\lambda)\mathbf{x} \in H$. Therefore a half space is convex. □

7.4
7.8
7.12
7.13
7.20

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and $-f$ also be convex. Observe that

$$\begin{aligned} f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &\leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) \\ \text{and} \quad -f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) &\leq -\lambda f(\mathbf{x}) - (1-\lambda)f(\mathbf{y}) \end{aligned}$$

Together, these two inequalities imply that

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) = \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y})$$

Thus f is linear, and given that any linear transformation is affine (see example 7.4.2), it follows that f is affine. □

7.21