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

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

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

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

□

(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) *Proof.* Let Q be positive definite. So, $f''(x) > 0$ for every x and Q is invertible. It follows that $\mathbf{x}^* = Q^{-1}\mathbf{b}$ for a given local minimizer $\mathbf{x}^*$. Now, let $\mathbf{x}^*$ be the unique minimizer of f. It follows that Q is positive semidefinite by the SONC. Given that $\mathbf{x}^*$ uniquely solves $Q^T\mathbf{x}^* = \mathbf{b}$, then Q is an invertible matrix (no eigenvalues which are 0). It follows that Q must be positive definite. □

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. □

6.15 See Jupyter notebook.

7.1

Proof. Let S be a nonempty subset of V. By definition, the *convex hull* of S, denoted 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 $\lambda_1 + ... + \lambda_k = 1$. So, let $y_1, y_2 \in \text{conv}(S)$. Hence there is some $k \in \mathbb{N}$ and $(\lambda)_{i=1}^{k}$ such that $y = \sum_{i=1}^{k} \lambda_i y_i$. Furthermore, there is some $j \in \mathbb{N}$ and $(\xi_i)_{i=1}^{j}$ such that $x = \sum_{i=1}^{j} \xi_i x_i$. Given that $\lambda \in [0,1]$, observe that

$$\lambda x + (1-\lambda)y = \lambda \sum_{i=1}^{j} \xi_i x_i + (1-\lambda) \sum_{i=1}^{k} \lambda_i y_i \quad \text{where}$$

$$\begin{aligned} \lambda\left(\sum_{i=1}^{j} \xi_i\right) + (1-\lambda)\sum_{i=1}^{k} \lambda_i &= \lambda + (1-\lambda) \\ &= 1 \end{aligned}$$

Hence $\lambda x + (1-\lambda)y$ is a convex combination of the elements of S. Thus $\lambda x + (1-\lambda)y \in \text{conv}(S)$ and $\text{conv}(S)$ is convex. □

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

(i) *Proof.* Let $x, y, p \in \mathbb{R}^n$. Observe that

$$\begin{aligned}
||x - y||^2 &= ||x - p + p - y||^2 \\
&= \langle x - p + p - y, x - p + p - y||^2\rangle \\
&= \langle x - p, x - p\rangle + \langle x - p, p - y\rangle + \langle p - y, x - p\rangle + \langle p - y, p - y\rangle \\
&= ||x - p||^2 + 2\langle x - p, p - y\rangle + ||p - y||^2
\end{aligned}$$

$\square$

(ii) *Proof.* Let $C \in \mathbb{R}^n$ be nonempty, closed, and convex. Suppose we have a point $p \in C$, and let $\langle x - p, p - y\rangle \geq 0$ for every $y \in C$. Observe that $||x - y||^2 = ||x - p||^2 + 2\langle x - p, p - y\rangle + ||p - y||^2$ by part (i). It follows that $||x - y||^2 \geq 0, ||x - p||^2 \geq 0$, and $||p - y||^2 \geq 0$ by definition. Now, $y \neq p$; hence, $p - y \neq 0$ and $||p - y||^2 > 0$. So, observe that

$$\begin{aligned}
||x - y||^2 &= ||x - p||^2 + 2\langle x - p, p - y\rangle + ||p - y||^2 \\
&\geq ||x - p||^2 + ||p - y||^2 \\
&> ||x - p||^2
\end{aligned}$$

$\square$

(iii) *Proof.* Let $z = \lambda y + (1 - \lambda)p$, where $0 \leq \lambda \leq 1$. By linearity, observe that

$$\begin{aligned}
||x - z||^2 &= ||x - \lambda y - (1 - \lambda)p||^2 \\
&= ||x - p + \lambda(p - y)||^2 \\
&= \langle x - p, x - p\rangle + 2\lambda\langle x - p, p - y\rangle + \lambda^2||p - y||^2 \\
&= ||x - p||^2 + 2\lambda\langle x - p, p - y\rangle + \lambda^2||y - p||^2
\end{aligned}$$

$\square$

(iv) *Proof.* Let p be a projection of x onto the convex set C. By part (iii), we have $||x - z||^2 = ||x - p||^2 + 2\lambda\langle x - p, p - y\rangle + \lambda^2||y - p||^2$. Given that $||x-z|^2 \geq 0$, $||x-p||^2 \geq 0$, and $\lambda \geq 0$, observe that

$$\begin{aligned} 0 \leq ||x-z||^2 &= ||x-p||^2 + 2\lambda\langle x-p, p-y\rangle + \lambda^2||y-p||^2 \\ &\leq 2\lambda\langle x-p, p-y\rangle + \lambda^2||y-p||^2 \end{aligned}$$

By letting $0 < \lambda \leq 1$, it follows that $0 \leq 2\langle x-p, p-y\rangle + \lambda||y-p||^2$. □

(v) Now, we prove the Theorem.

Proof. Let a point $p \in C$ be the projection of x onto C. By part (iv), choose $\lambda = 0$. Hence $\langle x-p, p-y\rangle \geq 0$ for every $y \in C$.
Now, let $\langle x-p, p-y\rangle \geq 0$ for every $y \in C$. By part (i), it follows that $||x-y|| > ||x-p||$ for all $y \in C$ and $y \neq p$. Therefore, $p \in C$ is the projection of x onto C. □

7.8

Proof. Let $f : \mathbb{R}^m \to \mathbb{R}$ be convex, $A \in M_{m\times n}\mathbb{R}$, and $b \in \mathbb{R}^m$. Now, choose some $x_1, x_2 \in \mathbb{R}^m$, and fix $\lambda \in [0,1]$. Observe that

$$\begin{aligned} g(\lambda x_1 + (1-\lambda)x_2) &= f(A(\lambda x_1 + (1-\lambda)x_2 + b) \\ &= f(\lambda(Ax_1 + b) + (1-\lambda)(Ax_2 + b)) \\ &\leq \lambda f(Ax_1 + b) + (1-\lambda)f(Ax_2 + b) \\ &= \lambda g(x_1) + (1-\lambda)g(x_2) \end{aligned}$$

Therefore, $g(x) = f(Ax+b)$ is convex. □

7.12

(i) *Proof.* Let $PD_n(\mathbb{R})$ be a set of positive-definite matrices in $M_n\mathbb{R}$). Now, choose two positive-definite matrices Y, Z and let $\lambda \in [0,1]$ with $x \in \mathbb{R}^n$. So, $x^TYx > 0$ and x^TZx by definition. Observe that for $\lambda Y + (1-\lambda)Z$ we have that

$$x^T(\lambda Y + (1-\lambda)Z)x = \lambda x^TYx + (1-\lambda)x^TZx$$

Now, if it's the case that $0 < \lambda < 1$, then the equality above is strictly positive. Similarly, if we look at the boundary, then the entire thing is strictly positive. Therefore, it follows that $\lambda Y + (1-\lambda)Z$ is positive-definite. Therefore $PD_n(\mathbb{R})$ is convex. $\square$

(ii) *Proof.* Let $f(X) = -\log(\det(X))$

(a) Let $A, B \in PD_n(\mathbb{R})$ such that $g(t) : [0,1] \to \mathbb{R}$ is given by $g(t) = f(tA + (1-t)B)$. Observe that

$$\begin{aligned} f(tA + (1-t)B) &= g(t) \\ &= g(1 \cdot t + (1-t) \cdot 0) \\ &\leq tg(1) + (1-t)g(0) \\ &= tf(A) + (1-t1)f(B) \end{aligned}$$

and f is convex.

(b) Now, given that positive definite matrices are normal, (call such a given matrix, A) we know there exists a nonsingular matrix S where $A = S^H S$. Observe that

$$\begin{aligned} -\log(\det(S^H(tI + (1-t)(S^H)^{-1}BS^{-1})S)) &= \\ -\log(\det(S^H)\det(tI + (1-t)(S^H)^{-1}BS^{-1})\det(S)) &= \\ -\log(\det(S^H))\det(S)\det(tI + (1-t)(S^H)^{-1}BS^{-1}) &= \\ -\log(\det(S^H S)\det(tI + (1-t)(S^H)^{-1}BS^{-1})) &= \\ -\log(\det(S^H S) - \log\det(tI + (1-t)(S^H)^{-1}BS^{-1})) &= \\ -\log(\det(A) - \log\det(tI + (1-t)(S^H)^{-1}BS^{-1})) & \end{aligned}$$

(c) Now, suppose $\lambda_1, ..., \lambda_n$ are the eigenvalues of $(S^H)^{-1}BS{-}1$. Note that $\{t+(1-t)\lambda\}_{i=1}^n$ are the eigenvalues of $(tI+(1-t)(S^H)^{-1}BS^{-1})$. Observe that

$$\begin{aligned} (tI + (1-t)(S^H)^{-1}BS^{-1})x_i &= tx_i + (1-t)\lambda_i x_i \\ &= (t + (1-t)\lambda_i)x_i \end{aligned}$$

Thus,

$$
\begin{aligned}
g(t) &= \log(\det(A) - \log\det(tI + (1-t)(S^H)^{-1}BS^{-1})) \\
&= -\sum_{i=1}^{n} \log(t + (1-t)\lambda_i) - \log(\det(A))
\end{aligned}
$$

(d) Using part (c), observe that

$$
\begin{aligned}
g'(t) &= -\sum_{i=1}^{n} \frac{(1-\lambda_i)}{\log(t + (1-t)\lambda_i)} \\
g''(t) &= \sum_{i=1}^{n} \frac{(1-\lambda_i)^2}{\log(t + (1-t)\lambda_i)^2} \geq 0 \;\; \forall t \in [0,1]
\end{aligned}
$$

Therefore, the function $f(X) = -\log(\det(X))$ is convex on $PD_n(\mathbb{R})$ $\square$

7.13

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and bounded above, and suppose by way of contradiction that f is not constant. So, there are points $x_1, x_2 \in \mathbb{R}^n$ such that $f(x_1) \neq f(x_2)$. Now, assume $f(x) < M$ for every x. Given some $\lambda \in [0,1]$, observe that

$$
\begin{aligned}
f(x_1) &= f\left(\lambda \frac{x_1 - (1-\lambda)x_2}{\lambda} + (1-\lambda)x_2\right) \\
&\leq \lambda f\left(\frac{x_1 - (1-\lambda)x_2}{\lambda}\right) + (1-\lambda)f(x_2)
\end{aligned}
$$

Hence, $\frac{f(x_1)-(1-\lambda)f(x_2)}{\lambda} \leq f\left(\frac{x_1-(1-\lambda)x_2}{\lambda}\right) \leq M$. Thus $\frac{f(x_1)-f(x_2)}{\lambda} \leq \frac{f(x_1)-(1-\lambda)f(x_2)}{\lambda}$. So $\frac{f(x_1)-f(x_2)}{\lambda} \leq M$. Note that as $\lambda \to 0$, the left-hand side becomes unbounded which is a contradiction. Therefore, f is constant. $\square$

7.20

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

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

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

Proof. Let $D \subset \mathbb{R}$ with $f : \mathbb{R}^n \to D$, and let $\phi : D \to \mathbb{R}$ be a strictly increasing function.
Now, suppose x^* is a local minimizer of f. It follows that there is some neighborhood denoted Ω such that

$$f(x^*) \le f(x) \ \ \forall x \in \Omega$$

Given that ϕ is strictly increasing, it follows that

$$\phi(f(x^*)) \le \phi(f(x)) \ \ \forall x \in \Omega$$

and x^* is a local minimizer of $\phi(f(x))$.
Now, let x^* be a local minimizer of $\phi(f(x))$. It follows that there is some neighborhood denoted Γ such that

$$\phi(f(x^*)) \le \phi(f(x)) \ \ \forall x \in \Gamma$$

Given that ϕ is strictly increasing, it must have an inverse (since it is injective). Hence, ϕ^{-1} must also be strictly increasing. So, we have that

$$\phi^{-1}\phi(f(x^*)) \le \phi^{-1}\phi(f(x)) \ \ \forall x \in \Gamma \tag{0.1}$$

So, it follows that $f(x^*) = f(x)$ for every $x \in \Gamma$ and x^* is a local minimizer of f. $\square$