1. 3.2.1
   In order for
   $$\frac{|x^\top y|^p}{\|x\| + \|y\|}$$
   to be differentiable, we would first need it to be continuous at 0. This means that
   $$\lim_{(x,y)\to(0,0)} \frac{|x^\top y|^P}{\|x\| + \|y\|} = 0$$
   We can make the substitution $x = r \cdot u, y = s \cdot v$ where u, v are unit vectors and $r = \|x\|, s = \|y\|$. This means that we get
   $$\lim_{r,s\to 0} \frac{|ru^T \cdot sv|^P}{r + s} = \frac{r^P s^P}{r + s} |u^T v| = 0$$
   which is true if $\lim_{r,s\to 0} \frac{r^P s^P}{r+s} = 0$. Since we are interested in (x, y) near $(0, 0)$, we can let $x = \varepsilon \cdot u, y = \varepsilon \cdot v$ for ε small. Thus, we get
   $$\lim_{r,s\to 0} \frac{r^P s^P}{r + s} = \lim_{\varepsilon\to 0} \frac{\varepsilon^{2P}}{2\varepsilon}$$
   $$= \frac{\varepsilon^{2P-1}}{2}$$
   If $p > \frac{1}{2}$, we get that $f(x, y) \to 0$ as $\varepsilon \to 0$. Thus, the function is continuous when $p > \frac{1}{2}$. If $p \le \frac{1}{2}$, the function doesn't approach 0 and hence the isn't continuous at $(0, 0)$. Thus, we must only analyze the differentiablity of $f(x, y)$ when $p > \frac{1}{2}$.
   First we will consider the partial derivatives.

   The partial derivatives at $(0, 0)$ are:
   $$\frac{\partial f}{\partial x_i}(0, 0) = \lim_{t\to 0} \frac{f(te_i, 0) - f(0, 0)}{t} = 0,$$
   $$\frac{\partial f}{\partial y_i}(0, 0) = \lim_{t\to 0} \frac{f(0, te_i) - f(0, 0)}{t} = 0.$$
   So, the gradient $\nabla f(0, 0) = 0$

   Clearly, since the partials are continuous, the function is also Frechet differentiable. But we must look closer at the values of P. Looking at the limit, we get
   $$\lim_{(x,y)\to(0,0)} \frac{f(x, y) - f(0, 0) - \nabla f(0, 0) \cdot (x, y)}{\|(x, y)\|} = 0$$
   This simplifies to
   $$\lim_{(x,y)\to(0,0)} \frac{f(x, y)}{\|(x, y)\|} = 0$$
   Using the substitution we did above, we get
   $$\frac{f(h, k)}{\|(h, k)\|} = \frac{\epsilon^{2p-1}}{2\sqrt{2}\epsilon} |u^\top v|^p = \frac{\epsilon^{2p-2}}{2\sqrt{2}} |u^\top v|^p$$

Hence, for $p > 1$, we see that $2p - 2 > 0$ and thus the limit is 0. Thus, the function is differentiable. If $p = 1$, we get $2p - 2 = 0$ so the limit won't be 0 and hence the function won't be differentiable. So it is only Frechet differentiable if $p > 1$.

2. 3.2.2

Let f be Lipschitz and $f(x_0) = 0$. Now let $g(x) = (f(x))^2$. Consider the limit

$$\lim_{x \to x_0} \frac{\|g(x) - g(x_0) - g'(x_0)(x - x_0)\|}{\|x - x_0\|}$$

This can be simplified to

$$\lim_{x \to x_0} \frac{\|g(x) - g'(x_0)(x - x_0)\|}{\|x - x_0\|}$$

Now, looking at the derivative of g, we see that $\nabla g(x) = 2 \cdot f(x) \cdot \nabla f(x)$ and hence $\nabla g(x_0) = 2 \cdot f(x_0) \cdot \nabla f(x_0) = 0$. Thus, we are just left with

$$\lim_{x \to x_0} \frac{\|(f(x))^2\|}{\|x - x_0\|}$$

We can rewrite this as

$$\lim_{x \to x_0} \frac{\|(f(x))^2\|}{\|x - x_0\|} = \lim_{h \to 0} \frac{\|(f(x_0 + h))^2\|}{\|h\|}$$

By the Lipschitz condition, we get that $|f(x_0 + h) - f(x_0)| = |f(x_0 + h)| \le L\|h\|$. Thus, $f(x_0 + h)^2 \le L^2\|h\|^2$. Hence, the limit simplifies to

$$\begin{aligned}
\lim_{h \to 0} \frac{\|(f(x_0 + h))^2\|}{\|h\|} &\le \lim_{h \to 0} \frac{L^2\|h\|^2}{\|h\|} \\
&= \lim_{h \to 0} L^2\|h\| \\
&= 0
\end{aligned}$$

Thus, the function is Frechet differentiable.

3. 3.2.3

Define $g(\lambda) = f(\lambda x + (1 - \lambda)y)$. Thus, $g(1) = f(x)$ and $g(0) = f(y)$. We can compute $g'(c)$ and get $\nabla f(x)(y - x)$. Now, since f is convex, so is g. This means we can use the tangent line inequality because g is convex and see that $g(1) \ge g(0) + g'(0)(1 - 0)$. The tangent line approximation of $f(y)$ at x is $T = f(x) + \nabla f(x)(y - x)$. But we know that f is convex so that means the function is above its tangent lines. This means that $f(y) \ge f(x) + \nabla f(x)(y - x)$

4. 3.2.4

Simple computation shows that $\nabla f(0, 0) = (1, 0)^T$ since $f_x(0, 0) = 1, f_y(0, 0) = 0$. Thus,

the linear map would be just $f'(0,0)((x,y) - (0,0)) = x$. To evaluate Frechet differentiablity, we need to check the limit

$$\lim_{(x,y) \to (0,0)} \frac{f(x,y) - f(0,0) - x}{\sqrt{x^2 + y^2}}$$

But we can simplify this to be $x \left(\frac{\sin(y)}{y} - 1 \right)$ when $y \neq 0$. But we can approximate $\frac{\sin(y)}{y} - 1$ by $\frac{y^2}{6}$ using the taylor expansion. Thus, we approximate $x \left(\frac{\sin(y)}{y} - 1 \right)$ by $\frac{xy^2}{6}$. Hence, the expression we take the limit of becomes

$$\frac{\|xy^2\|}{6\sqrt{x^2 + y^2}}$$

Taking the limit as $(x,y) \to (0,0)$, we get 0 regardless of the path. Hence, the function is Frechet differentiable.

5. 3.3.1

This is a direct application of the Inverse Function Theorem. Since f is C^1 and has invertible derivative matrix on any x, we know that f has invertible derivative matrix on any open $G \subseteq \mathbb{R}^m$. Thus, by the Inverse function theorem, we know that $f(G)$ is open.

6. 3.3.2

Define $F(x,y) = xf(y) + g(x)y + h(x)$. Then it is clear that $F(1,1) = 1 \cdot f(1) + g(1) \cdot 1 + h(1) = f(1) + g(1) + h(1) = 0$. This suggests that $y = 1$ is a solution of the equation when $x = 1$. Computing the partial derivative of F, we see that $\frac{\partial F}{\partial y} = xf'(y) + g(x)$. At $(1,1)$, we get that $\frac{\partial F}{\partial y}(1,1) = 1 \cdot f'(1) + g(1) = f'(1) + g(1) > 0$ by the assumption that $g > 0$ and f is non decreasing. Since the derivative is nonzero, we can apply the Implicit function theorem and state that there is a $\varphi(x) \in C^1$ on a neighborhood of x such that $F(x, \varphi(x)) = 0$ for all $x \in (1 - \delta, 1 + \delta)$. Thus $F(x, \varphi(x)) = xf(\varphi(x)) + g(x)\varphi(x) + h(x) = 0$.

7. 3.3.3

Since $\phi_x(\bar{x}) \neq 0$, there is $k \in \{1, ..., m\}$ such that $\frac{\partial f}{\partial x_k} \neq 0$. We now apply the implict function theorem and state that there is a neighborhood U of $\bar{x} \subseteq \mathbb{R}^m$ and a neighborhood V of $(\bar{x}_1, ..., \bar{x}_{k-1}, \bar{x}_{k+1}, ..., \bar{x}_m) \subseteq \mathbb{R}^{m-1}$ and a unique function C^1 function θ such that

$$\phi(x_1, ..., x_{k-1}, \theta(x_1, ..., x_{k-1}, x_{k+1}, ..., x_m), x_{k+1}, ..., x_m) = 0$$

To find δ and Γ, we choose $\delta > 0$ be small enough such that $B(0, \delta) + \bar{x} \subseteq U$ and define $\Gamma = \{(x_1, ..., x_{k-1}, x_{k+1}, ..., x_m) : (x_1, ..., x_{k-1}, x_{k+1}, ..., x_m) \in B(0, \delta) + \bar{x}\}$. The geometric interpratation would be that near any boundary point of G, the boundary can be locally represented as the $m-1$ dimensional surface in $\mathbb{R}^m$ and that locally, that surface behaves like θ, which is smooth. So locally the boundary is smooth.

8. 3.4.1

Let g be the interpolation function. In order for g to be smooth, we need it to have derivatives of all order. In order to interpolate smoothly, we need $g(1) = 1, g(2) = 0, g'(1) = g'(2) = 0$. One option would be to try the cubic function $g(x) = ax^3 + bx^2 + cx + d$ and solve the system of equations

$$a + b + c + d = 1,$$
$$8a + 4b + 2c + d = 0,$$
$$3a + 2b + c = 0,$$
$$12a + 4b + c = 0.$$

in order to find a, b, c, d. Doing so gives $a = 2$ and $b = -9$ and $c = 12$ and $d = -4$. Thus, the cubic function $g(x) = 2x^3 - 9x^2 + 12x - 4$ smoothly interpolates between the two as desired.

9. 3.4.2

Let the Hessian of f be positive semi definate. This means that $v^T f_{xx}(x)v \geq 0$ for all $v \in \mathbb{R}^M$. We take the taylor expansion of f around x. For any $y \in \mathbb{R}^M$, we exapnd $f(y)$ around x as

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T f_{xx}(x)(y - x) + o(\|y - x\|^2)$$

Since $f_{xx}(x)$ is positive semi definate, we know that $(y - x)^T f_{xx}(x)(y - x) \geq 0$ and hence we get

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

But we know that this implies that f is convex.

10. 3.4.3

Define $g(t) = f(t) - yt$ for $y \in [f'(a), f'(b)]$. Since f is differentiable, then so is g and $g'(t) = f'(t) - y$. We can apply the mean value theorem and state that there is a $c \in [a, b]$ such that $\frac{f(b) - yb - (f(a) - ya)}{b - a} = g'(c)$. We can simplify that ratio to get

$$\frac{f(b) - f(a) - y(b - a)}{b - a} = \frac{f(b) - f(a)}{b - a} - y$$

This implies that $g'(c) = \frac{f(b) - f(a)}{b - a} - y = f'(c) - y$ and hence $f'(c) = \frac{f(b) - f(a)}{b - a}$. Thus, $f'(x)$ takes values between $f'(a)$ and $f'(b)$. Thus, by the intermediate value property, we can say that there is a c_1 such that $f'(c_1) = y$

11. 3.4.4

First, we take the limit as $x \to y$ and see that $\lim_{x \to y} |f(x) - f(y)| \leq \lim_{x \to y} K|x - y|^\alpha = 0$. This means that $f(y) = \lim_{x \to y} f(x)$ and hence f is continuous. Now, fix $y = 0$. Then we get $\|f(x) - f()\| \leq K\|x - 0\|^\alpha = K\|x\|^\alpha$. Since this must be true for all x, we know that we can make the right arbitrarily small by choosing a small x. Thus, the only way this is going to be true is if $f(x) = f(0)$.

12. 3.4.6

   (a) First note that f is a polynomial and hence differentiable. Let n be even. Then we know that $n-1$ is odd. Let $f(x) = x^n + ax + b$. Then $f'(x) = nx^{n-1} + a$ and since x^{n-1} is odd, the derivative will change sign as most once and thus, we know that f has at most one local extrema. Thus, it can only cross the x axis twice at most since the x^n term will dominate for large x and the function will be far from 0. Hence, the function can only have at most two real zeros.

   (b) First note that f is a polynomial and hence differentiable. By the fundamental theorem of algebra, we know that f has n roots. Since complex roots come in pairs, we know that there must be at least one real root. Call that root c. We can divide f by $(x-c)$ and get a new polynomial of degree $n-1$. Let $m = n-1$. Then m is even since n is even. Thus we get a polymial of the form $x^m + a$. Using the result from part 1, we know that this polynomial can have at most 2 real roots. Thus, f has at most 3 real roots.

13. 3.4.7

   Fix an $n \in \mathbb{N}$. Define $f(x) = a_0 x + \frac{a_1}{2}x^2 + \cdots + \frac{a_{n-1}}{n}x^n + \frac{a_n}{n+1}x^{n+1}$. It is clear that $f(0) = 0$ and $f(1) = a_0 + \frac{a_1}{2} + \cdots + \frac{a_{n-1}}{n} + \frac{a_n}{n+1} = 0$ as given in the problem. So $f(0) = f(1) = 0$. Since f is a polynomial, it is continuous and differentiable over $\mathbb{R}$. Thus, we can apply Rolle's Theorem and state that there is a point $c \in (0,1)$ such that $f'(c) = 0$. But $f'(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n$. Hence we know that $a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n = 0$ has a solution in $(0,1)$.

14. 3.4.9

   Using the taylor theorem, we can expand $f(x + 2h)$ around x and get $f(x + 2h) = f(x) + f'(x) \cdot 2h + 4h^2 f''(\eta)$ for some $\eta \in (x, x+2h)$. Now we subtract $f(x)$ from both sides and divide by $2h$ and get $\frac{f(x+2h)-f(x)}{2h} = f'(x) + hf''(\eta)$. Thus, we get

   $$f'(x) = \frac{f(x+2h) - f(x)}{2h} - hf''(\eta)$$

   Now, to handle the Supremum identity, we first bound the difference and see that

   $$\left| \frac{f(x+2h) - f(x)}{2h} \right| \leq \frac{2\sup_{x\in[a,b]} \|f(x)\|}{2h} = \frac{\sup_{x\in[a,b]} \|f(x)\|}{h}$$

   and we know that the supremum is well defined and finite since f is continuous. Similarly, we bound the second derivative term and see that

   $$\|hf''(\eta)\| \leq h \sup_{x\in[a,b]} \|f''(x)\|$$

   Now, we can combine the bounds and see that

   $$\|f'(x)\| \leq \frac{\sup_{x\in[a,b]} \|f(x)\|}{h} + h \sup_{x\in[a,b]} \|f''(x)\|$$

We can now better select the h such that the bound is sharpened. Thus, we need $h \sup_{x \in [a,b]} \|f''(x)\| = \frac{\sup_{x \in [a,b]} \|f(x)\|}{h}$. Now, we can solve for h and get

$$h = \sqrt{\frac{\sup_{x \in [a,b]} \|f(x)\|}{\sup_{x \in [a,b]} \|f''(x)\|}}$$

Substituting that back in, we get

$$\|f'(x)\| \leq 2 \sqrt{\sup_{x \in [a,b]} \|f''(x)\| \cdot \sup_{x \in [a,b]} \|f(x)\|}$$

Squaring both sides to remove the square root and taking the supremum, we get

$$\sup_{x \in [a,b]} \|f'(x)\|^2 \leq 4 \sup_{x \in [a,b]} \|f''(x)\| \cdot \sup_{x \in [a,b]} \|f(x)\|$$

15. 3.4.10

Since $f(0) = 0$, so will be all its derivatives. That is, $f^n(0) = 0$ for all n where f^i denotes the ith derivative. Thus, the taylor expansion of f at $x = 0$ would be 0. But clearly, f is not equal to 0 everywhere. Thus, the function is not equal to it's taylor expansion at 0. So even though it is smooth, it is not equal to its taylor expansion and thus we know that simply being smooth is not enough to be equal to its taylor expansion.