Exercise 8.1

We have

$$\begin{aligned} f(x,y)-f(x_0,y_0) &= a(x_0+\Delta x)^2+2b(x_0+\Delta x)(y_0+\Delta y)+c(y_0+\Delta y)^2-ax_0^2-2bx_0y_0-cy_0^2 \\ &= (2ax_0+2by_0)\Delta x+(2bx_0+2cy_0)\Delta y+(a\Delta x^2+2b\Delta x\Delta y+c\Delta y^2). \end{aligned}$$

By

$$|a\Delta x^2+2b\Delta x\Delta y+c\Delta y^2| \le (|a|+2|b|+|c|)\max\{|\Delta x|,|\Delta y|\}^2 \le (|a|+2|b|+|c|)\epsilon\max\{|\Delta x|,|\Delta y|\},$$

the function is differentiable, with

$$f'(x_0,y_0)(u,v)=(2ax_0+2by_0)u+(2bx_0+2cy_0)v.$$

Exercise 8.2

Suppose $F(\vec{x})$ is differentiable at $\vec{x}_0$. For any $\epsilon>0$, there is $\delta>0$, such that

$$\|\vec{x}-\vec{x}_0\|<\delta \implies \|F(\vec{x})-\vec{a}-L(\Delta\vec{x})\|\le\epsilon\|\Delta\vec{x}\|.$$

Then

$$\vec{a}=F(\vec{x}_0),\quad \|F(\vec{x})-F(\vec{x}_0)\|\le\|L(\Delta\vec{x})\|+\epsilon\|\Delta\vec{x}\|\le(\|L\|+\epsilon)\|\Delta\vec{x}\|<(\|L\|+\epsilon)\delta.$$

This shows that

$$\|\Delta\vec{x}\|<\min\left\{\delta,\frac{\epsilon}{\|L\|+\epsilon}\right\} \implies \|F(\vec{x})-F(\vec{x}_0)\|<\epsilon.$$

This proves that $F(\vec{x})$ is continuous at $\vec{x}_0$.

By Exercise 6.21, the norm is always continuous. Suppose $\|\vec{x}\|_2$ has linear approximation $a+l(\vec{x})$ at $\vec{0}$. Then $a=\|\vec{0}\|_2=0$ and for any $\epsilon>0$, there is $\delta>0$, such that

$$\|\vec{x}\|_2<\delta \implies |\|\vec{x}\|_2-l(\vec{x})|\le\epsilon\|\vec{x}\|_2.$$

Applying the property to $-\vec{x}$ and using the linearity of l, we find

$$\|\vec{x}\|_2<\delta \implies |\|\vec{x}\|_2+l(\vec{x})|\le\epsilon\|\vec{x}\|_2.$$

Therefore

$$\|\vec{x}\|_2<\delta \implies 2|l(\vec{x})|\le|\|\vec{x}\|_2-l(\vec{x})|+|\|\vec{x}\|_2+l(\vec{x})|\le 2\epsilon\|\vec{x}\|_2.$$

This further implies $|l(c\vec{x})|\le\epsilon\|c\vec{x}\|_2$, where $c\vec{x}$ can be any vector of arbitrary length. Since ϵ is arbitrary, we conclude that $l(\vec{x})=0$ for all $\vec{x}$. Thus for any $\epsilon>0$, there is $\delta>0$, such that $\|\vec{x}\|_2<\delta$ implies $|\|\vec{x}\|_2|\le\epsilon\|\vec{x}\|_2$. This cannot hold for $\epsilon<1$. The contradiction shows that $\|\vec{x}\|_2$ is not differentiable at $\vec{0}$.

Exercise 8.3

Denote $L = F'(\vec{x}_0)$. For any $\epsilon > 0$, there is $\delta > 0$, such that

$$\|\vec{x}-\vec{x}_0\| < \delta \implies \|F(\vec{x}) - F(\vec{x}_0) - L(\Delta x)\| = \|F(\vec{x}) - L(\Delta x)\| \le \epsilon\|\Delta x\|,\ |\lambda(\vec{x}) - \lambda(\vec{x}_0)| < \epsilon.$$

Then

$$\begin{aligned}
\|F(\vec{x})\| &\le \|L(\Delta x)\| + \epsilon\|\Delta x\| \le (\|L\| + \epsilon)\|\Delta x\|, \\
\|\lambda(\vec{x})F(\vec{x}) - \lambda(\vec{x}_0)L(\Delta x)\| &\le \|\lambda(\vec{x})F(\vec{x}) - \lambda(\vec{x}_0)F(\vec{x})\| + \|\lambda(\vec{x}_0)F(\vec{x}) - \lambda(\vec{x}_0)L(\Delta x)\| \\
&\le |\lambda(\vec{x}) - \lambda(\vec{x}_0)|\|F(\vec{x})\| + |\lambda(\vec{x}_0)|\|F(\vec{x}) - L(\Delta x)\| \\
&\le \epsilon(\|L\| + \epsilon)\|\Delta x\| + \epsilon|\lambda(\vec{x}_0)|\|\Delta x\| \\
&= \epsilon(\|L\| + \epsilon + |\lambda(\vec{x}_0)|)\|\Delta x\|.
\end{aligned}$$

This proves that $\lambda(\vec{x})F(\vec{x})$ is differentiable with $(\lambda F)'(\vec{x}_0) = \lambda(\vec{x}_0)L = \lambda(\vec{x}_0)F'(\vec{x}_0)$.

EXERCISE 8.4

If $f(\vec{x}) = f(\vec{x}_0) + J(\vec{x})\cdot\Delta\vec{x}$ and J is continuous, then

$$f(\vec{x}) = f(\vec{x}_0) + J(\vec{x}_0)\cdot\Delta\vec{x} + R(\vec{x})$$

with

$$|R(\vec{x})| = |(J(\vec{x}_0) - J(\vec{x}_0))\cdot\Delta\vec{x}| \le \|J(\vec{x}) - J(\vec{x}_0)\|_2\|\Delta\vec{x}\|_2.$$

The continuity of J at $\vec{x}_0$ then tells us $R(\vec{x}) = o(\|\Delta\vec{x}\|)$. Since $J(\vec{x}_0)\cdot\Delta\vec{x}$ is linear in $\Delta\vec{x}$, we conclude that f is differentiable at $\vec{x}_0$ and $\nabla f(\vec{x}_0) = J(\vec{x}_0)$.

Conversely, suppose f is differentiable at $\vec{x}_0$. Let

$$R(\vec{x}) = f(\vec{x}) - f(\vec{x}_0) - f'(\vec{x}_0)(\Delta\vec{x}) = f(\vec{x}) - f(\vec{x}_0) - \nabla f(\vec{x}_0)\cdot\Delta\vec{x}.$$

Then

$$\lim_{\Delta\vec{x}\to\vec{0}}\left\|\frac{R(\vec{x})\Delta\vec{x}}{\|\Delta\vec{x}\|_2^2}\right\| = \lim_{\Delta\vec{x}\to\vec{0}}\frac{|R(\vec{x})|}{\|\Delta\vec{x}\|_2} = 0.$$

Thus $J(\vec{x}) = \nabla f(\vec{x}_0) + \dfrac{R(\vec{x})\Delta\vec{x}}{\|\Delta\vec{x}\|_2^2}$ is continuous at $\vec{x}_0$, and

$$\begin{aligned}
f(\vec{x}_0) + J(\vec{x})\cdot\Delta\vec{x} &= f(\vec{x}_0) + \nabla f(\vec{x}_0)\cdot\Delta\vec{x} + \frac{R(\vec{x})\Delta\vec{x}}{\|\Delta\vec{x}\|_2^2}\cdot\Delta\vec{x} \\
&= f(\vec{x}_0) + \nabla f(\vec{x}_0)\cdot\Delta\vec{x} + R(\vec{x}) = f(\vec{x}).
\end{aligned}$$

EXERCISE 8.5

Suppose $f(c\vec{x}) = c^p f(\vec{x})$ for $c > 0$. If f is differentiable at $\vec{0}$. Then by restricting the definition to the straight line $t\vec{x}$ for fixed $\vec{x}$ and changing t, $f(t\vec{x})$ is differentiable at $t = 0$. This implies $f(t\vec{x}) = t^p f(\vec{x})$ is differentiable at $t = 0^+$. Thus we conclude that either $f(\vec{x}) = 0$ for all $\vec{x}$ or $p \ge 1$. The continuity at $\vec{0}$ then implies that $f(\vec{0}) = \vec{0}$.

Suppose $p > 1$. Then by $\vec{x} = r\vec{u}$, $r = \|\vec{x}\|$, $\|\vec{u}\| = 1$, we have $f(\vec{x}) = r^p f(\vec{u})$. If $|f(\vec{u})| < B$ is bounded on the unit sphere (i.e., for $\|\vec{u}\| = 1$), then $\|\vec{x}\| < \delta$ implies that $|f(\vec{x})| \le \delta^{p-1}B\|\vec{x}\|$, so that F is differentiable at $\vec{0}$ with $f'(\vec{0}) = 0$. If $f(\vec{u})$ is unbounded on the unit sphere, then no matter how small r is, we can always find $\vec{u}$ such that $|f(\vec{x})|$ is arbitrarily large. Thus the function is not continuous at $\vec{0}$. Consequently, the function is not differentiable at $\vec{0}$.

Suppose $p = 1$ and $f'(\vec{0}) = l$. Then $g(\vec{x}) = f(\vec{x}) - l(\vec{x})$ satisfies $g(c\vec{x}) = cg(\vec{x})$ for $c > 0$ and $g'(\vec{x}_0) = 0$. Thus for any $\epsilon > 0$, there is $\delta > 0$, such that $\|\vec{x}\| < \delta$ implies $|g(\vec{x})| \le \epsilon\|\vec{x}\|$. This further implies that $|g(c\vec{x})| = |cg(\vec{x})| \le \epsilon\|c\vec{x}\|$ for any $c > 0$. Since $c\vec{x}$ can be any nonzero vector of arbitrary length, we conclude that $|g(\vec{x})| \le \epsilon\|\vec{x}\|$ for all $\vec{x}$. Since ϵ is arbitrary, we have $g(\vec{x}) = 0$ for all $\vec{x}$. Thus $f(\vec{x}) = l(\vec{x})$ is a linear functional.

Exercise 8.6

We have

$$A\vec{x}\cdot\vec{x} - A\vec{x}_0\cdot\vec{x}_0 = A\Delta x\cdot\vec{x}_0 + A\vec{x}_0\cdot\Delta x + A\Delta x\cdot\Delta x.$$

Since $A\Delta x\cdot\vec{x}_0 + A\vec{x}_0\cdot\Delta x = (A + A^T)\vec{x}_0\cdot\Delta x$ is a linear functional of Δx and $\|A\Delta x\cdot\Delta x\|_2 \le \|A\|\|\Delta x\|_2^2 \le \epsilon\|A\|\|\Delta x\|_2$, we conclude that $A\vec{x}\cdot\vec{x}$ is differentiable, with the derivative at $\vec{x}_0$ to be $(A + A^T)\vec{x}_0\cdot\vec{v}$.

Exercise 8.7

We have

$$B(\vec{x}_0 + \Delta\vec{x}, \vec{y}_0 + \Delta\vec{y}) = B(\vec{x}_0, \vec{y}_0) + B(\Delta\vec{x}, \vec{y}_0) + B(\vec{x}_0, \Delta\vec{y}) + B(\Delta\vec{x}, \Delta\vec{y})$$

The term $B(\vec{x}_0, \vec{y}_0)$ is constant. The terms $B(\Delta x, \vec{y}_0)$ and $B(\vec{x}_0, \Delta y)$ are linear in $(\Delta\vec{x}, \Delta\vec{y})$. The term $B(\Delta\vec{x}, \Delta\vec{y})$ satisfies

$$\|B(\Delta\vec{x}, \Delta\vec{y})\| \le \|B\|\|\Delta\vec{x}\|\|\Delta\vec{y}\| \le \|B\|\|(\Delta\vec{x}, \Delta\vec{y})\|^2,$$

where $\|(\Delta\vec{x}, \Delta\vec{y})\| = \max\{\|\Delta\vec{x}\|, \|\Delta\vec{y}\|\}$. Therefore the first order derivative is given by the two linear terms

$$B'(\vec{x}_0, \vec{y}_0)(\Delta\vec{x}, \Delta\vec{y}) = B(\Delta\vec{x}, \vec{y}_0) + B(\vec{x}_0, \Delta\vec{y}).$$

Exercise 8.8

Let

$$\begin{aligned} \phi(t) &= \phi(t_0) + \phi'(t_0)\Delta t + R_1, & R_1 &= o(\Delta t); \\ \psi(t) &= \psi(t_0) + \psi'(t_0)\Delta t + R_2, & R_2 &= o(\Delta t). \end{aligned}$$

Then

$$\begin{aligned} B(\phi(t), \psi(t)) &= B(\phi(t_0) + \phi'(t_0)\Delta t + R_1, \psi(t_0) + \psi'(t_0)\Delta t + R_2) \\ &= B(\phi(t_0), \psi(t_0)) + [B(\phi'(t_0), \psi(t_0)) + B(\phi(t_0), \psi'(t_0))]\Delta t + R, \\ R &= B(\phi(t_0), R_2) + B(R_1, \psi(t_0)) + [B(\phi'(t_0), R_2) + B(R_1, \psi'(t_0))]\Delta t \\ &\quad + B(\phi'(t_0), \psi'(t_0))\Delta t^2 + B(R_1, R_2). \end{aligned}$$

By $\|B(\vec{u}, \vec{v})\| \le \|B\|\|\vec{u}\|\|\vec{v}\|$, we have

$$\|B(\phi(t_0), R_2)\| \le \|B\|\|\phi(t_0)\|\|R_2\|, \quad \|B(\phi'(t_0), R_2)\| \le \|B\|\|\phi'(t_0)\|\|R_2\|,$$

$$\|B(\phi'(t_0),\psi'(t_0))\| \le \|B\|\|\phi'(t_0)\|\|\psi'(t_0)\|, \quad \|B(R_1,R_2)\| \le \|B\|\|R_1\|\|R_2\|.$$

Then by $R_1 = o(\Delta t)$ and $R_2 = o(\Delta t)$, we get $B(\phi(t_0), R_2) = o(\Delta t)$, $B(\phi'(t_0), R_2) = o(\Delta t)$, $B(\phi'(t_0),\psi'(t_0))\Delta t^2 = o(\Delta t)$, $B(R_1,R_2) = o(\Delta t)$ and similar estimations for the other terms in R. We conclude that $R = o(\Delta t)$, so that $B(\phi(t_0),\psi(t_0)) + [B(\phi'(t_0),\psi(t_0)) + B(\phi(t_0),\psi'(t_0))]\Delta t$ is linear approximation of $B(\phi(t),\psi(t))$ at t_0. In particular, we get

$$[B(\phi(t),\psi(t))]'(t_0) = B(\phi'(t_0),\psi(t_0)) + B(\phi(t_0),\psi'(t_0)).$$

Exercise 8.9

By b linear in the first variable, we have

$$\left.\frac{d}{dt}\right|_{t=t_0} b(\phi(t),\vec{w}) = \lim_{t\to t_0} \frac{b(\phi(t),\vec{w}) - b(\phi(t_0),\vec{w})}{t-t_0} = \lim_{t\to t_0} b\left(\frac{\phi(t)-\phi(t_0)}{t-t_0}, \vec{w}\right).$$

Since b is a dual pairing, by Exercise 7.35, the right side equals $b(\vec{v},\vec{w})$ for all $\vec{w}$ if and only if

$$\phi'(t_0) = \lim_{t\to t_0} \frac{\phi(t)-\phi(t_0)}{t-t_0} = \vec{v}.$$

Exercise 8.10

Let $H = X - A$ (this is ΔX). Then $F(X) - F(A) = (A+H)^T(A+H) - A^TA = A^TH + H^TA + H^TH$. Since $A^TH + H^TA$ is linear in H and $\|H^TH\| \le \|H^T\|\|H\| \le c\|H\|^2 \le c\epsilon\|H\|$ ($\|H^T\|$ is also a norm of H, so that $\|H^T\| \le c\|H\|$). Thus F is differentiable, with $F'(A)(H) = A^TH + H^TA$.

Exercise 8.11

Suppose $f(X)$ is function of matrix that is multilinear with respect to the columns of X. Then we have

$$\begin{aligned} f(A+H) &= f(\vec{a}_1+\vec{h}_1, \vec{a}_2+\vec{h}_2, \dots, \vec{a}_n+\vec{h}_n) \\ &= f(\vec{a}_1,\vec{a}_2,\dots,\vec{a}_n) \\ &\quad + \sum_{1\le i\le n} f(\vec{a}_1,\vec{a}_2,\dots,\vec{h}_i,\dots,\vec{a}_n) \\ &\quad + \sum_{1\le i<j\le n} f(\vec{a}_1,\vec{a}_2,\dots,\vec{h}_i,\dots,\vec{h}_j,\dots,\vec{a}_n) \\ &\quad + \cdots \\ &\quad + f(\vec{h}_1,\vec{h}_2,\dots,\vec{h}_n). \end{aligned}$$

The first order part gives us the derivative of f at A

$$f'(A)(H) = \sum_{1\le i\le n} f(\vec{a}_1,\vec{a}_2,\dots,\vec{h}_i,\dots,\vec{a}_n).$$

In case $f = \det$ and the rank of $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n$ is $\le n-2$, replacing one vector gives $\vec{a}_1, \vec{a}_2, \dots, \vec{h}_i, \dots, \vec{a}_n$, with rank at most $n-1$. Therefore all the determinants vanish, and $\det'(A)(H) = 0$.

EXERCISE 8.12

We have

$$(A+H)^k - A^k = A^{k-1}H + A^{k-2}HA + \cdots + AHA^{k-2} + AH^{k-1} + R,$$

where $A^{k-1}H + A^{k-2}HA + \cdots + AHA^{k-2} + AH^{k-1}$ is linear in H, and R is a finite sum of products of i copies of H and $k-i$ copies of A in all possible order, with $i \ge 2$. Because $i \ge 2$, the norm of each such product is $\le \|A\|^{k-i}\|H\|^i \le \|A\|^{k-i}\epsilon\|H\|$. Thus $A^{k-1}H + A^{k-2}HA + \cdots + AHA^{k-2} + AH^{k-1}$ is the derivative of X^k at A.

EXERCISE 8.13

For $F(X) = X^{-1}$, we have

$$\begin{aligned} F(I+H) &= I - H + H^2 - H^3 + \cdots \\ &= I - H + H^2(I - H + H^2 - H^3 + \cdots) \\ &= F(I) - H + H^2(I+H)^{-1}. \end{aligned}$$

By Exercise 7.17, for $\|H\| < 1$ we also have

$$\|(I+H)^{-1} - I\| \le \frac{\|H\|\|I\|^2}{1 - \|H\|\|I\|} = \frac{\|H\|}{1-\|H\|},$$

so that

$$\|(I+H)^{-1}\| \le \|I\| + \frac{\|H\|}{1-\|H\|} = \frac{1}{1-\|H\|}.$$

Thus

$$\|H^2(I+H)^{-1}\| \le \|H\|^2\|(I+H)^{-1}\| \le \frac{\|H\|^2}{1-\|H\|} \le \epsilon\|H\|.$$

This shows that the linear map $-H$ of H is the derivative of $F(X)$ at I.

EXERCISE 8.14 (1)

We have $f(x,0) = |x|^{2p}\sin\dfrac{1}{x^2}$. Thus $f_x(0,0)$ exists if and only if $2p > 1$. Conversely, for $p > \dfrac{1}{2}$, we have $f_x(0,0) = f_y(0,0) = 0$.

If the function is differentiable at $(0,0)$, then $p > \dfrac{1}{2}$ and the linear approximation is 0. This means that for any $\epsilon > 0$, there is $\delta > 0$, such that $\sqrt{x^2+y^2} < \delta$ implies $\left|(x^2+y^2)^p \sin\dfrac{1}{(x^2+y^2)}\right| < \epsilon\sqrt{x^2+y^2}$. Letting $r = \sqrt{x^2+y^2}$, we find $0 < r < \delta$ implies $\left|r^{2p}\sin\dfrac{1}{r^2}\right| \le \epsilon r$. Since $p > \dfrac{1}{2}$, this always happens. Thus f is differentiable at $(0,0)$ if and only if $p > \dfrac{1}{2}$.

Exercise 8.14 (2)

The function is continuous at $(0,0)$ if and only if $p>0$ or $q>0$.

When $p>0$, we have $f(0,y)=0$ for all y. The partial derivative $f_y(0,0)=0$ exists. When $p=0$, we have $f(0,y)=\begin{cases}|y|^q\sin\dfrac{1}{y^2} & \text{if } y\neq 0\\ 0 & \text{if } y=0\end{cases}$, and $f_y(0,y)$ exists (and must be zero) if and only if $q>1$. Similarly, $f_x(0,0)$ exists if and only if $q>0$ or $p>1$, $q=0$.

Suppose $p>0$ and $q>0$. If the function is differentiable at $(0,0)$, then by the computation of the partial derivatives, the linear approximation must be 0. Therefore we need $\lim_{(x,y)\to(0,0)}\dfrac{|x|^p|y|^q}{\max\{|x|,|y|\}}\sin\dfrac{1}{x^2+y^2}=0$. By restricting the limit to the line $|x|=|y|$, we get $p+q>1$.

Now assume $p>0$, $q>0$ and $p+q>1$, then $|f(x,y)|\le\|(x,y)\|_\infty^{p+q}\le\epsilon\|(x,y)\|_\infty$ when $\|(x,y)\|_\infty<\epsilon^{\frac{1}{p+q-1}}$. Therefore the function is differentiable at $(0,0)$ with 0 as the linear approximation.

Exercise 8.14 (3)

By $f(x,0)=f(0,y)=0$, the function has partial derivatives $f_x(0,0)=f_y(0,0)=0$.

If the function is differentiable at $(0,0)$, then the linear approximation is 0. This means that

$$\lim_{x,y\to 0^+}\frac{x^py^q}{(x^m+y^n)^k(x+y)}=0.$$

When restricted to $x^m=y^n$, we have

$$\frac{x^{p+q\frac{m}{n}}}{2^{k+1}x^{mk}x^{\frac{1}{n}\min\{m,n\}}}\le\frac{x^py^q}{(x^m+y^n)^k(x+y)}=\frac{x^{p+q\frac{m}{n}}}{2^kx^{mk}(x+x^{\frac{m}{n}})}\le\frac{x^{p+q\frac{m}{n}}}{2^kx^{mk}x^{\frac{1}{n}\min\{m,n\}}}.$$

Therefore the restriction converges to 0 if and only if $p+q\dfrac{m}{n}>mk+\dfrac{1}{n}\min\{m,n\}$. This is the same as

$$\frac{p}{m}+\frac{q}{n}>k+\frac{\min\{m,n\}}{mn}.$$

When restricted to $x=y$, we have

$$\frac{x^{p+q}}{2^{k+1}x^{\min\{m,n\}k}x}\le\frac{x^py^q}{(x^m+y^n)^k(x+y)}=\frac{x^{p+q}}{2(x^m+x^n)^kx}\le\frac{x^{p+q}}{2x^{\min\{m,n\}k}x}.$$

Therefore the restriction converges to 0 if and only if

$$p+q>\min\{m,n\}k+1.$$

Conversely, assume the two inequalities hold. Without loss of generality, assume $m\ge n$. Then the two inequalities mean that

$$\frac{p}{m}+\frac{q}{n}>k+\frac{1}{m},\quad p+q>nk+1.$$

Then we consider three regions.

For $y \le x^{\frac{m}{n}}$, we have

$$\frac{x^p y^q}{(x^m+y^n)^k(x+y)} \le \frac{x^{p+q\frac{m}{n}}}{x^{mk}x} = x^{p+q\frac{m}{n}-mk-1}.$$

By $p+q\dfrac{m}{n}-mk-1 = m\left(\dfrac{p}{m}+\dfrac{q}{n}-k-\dfrac{1}{m}\right) > 0$, the restriction of the limit on the region converges to 0.

For $x^{\frac{m}{n}} \le y \le x$, we have

$$\frac{x^p y^q}{2^k y^{nk} 2x} \le \frac{x^p y^q}{(x^m+y^n)^k(x+y)} \le \frac{x^p y^q}{y^{nk}x}.$$

Therefore $\lim_{x^{\frac{m}{n}} \le y \le x; x,y\to 0^+} \dfrac{x^p y^q}{(x^m+y^n)^k(x+y)} = 0$ if and only if $\lim_{x^{\frac{m}{n}} \le y \le x; x,y\to 0^+} x^{p-1}y^{q-nk} = 0$. For the given x, the maximum and the minimum of $x^{p-1}y^{q-nk}$ for the y in the range are $x^{p-1}x^{q-nk} = x^{p+q-nk-1}$ and $x^{p-1}x^{\frac{m}{n}(q-nk)} = x^{m\left(\frac{p}{m}+\frac{q}{n}-k-\frac{1}{m}\right)}$. Therefore $\lim_{x^{\frac{m}{n}} \le y \le x; x,y\to 0^+} x^{p-1}y^{q-nk} = 0$ if and only if $\lim_{x\to 0^+} x^{p+q-nk-1} = 0$ and $\lim_{x\to 0^+} x^{m\left(\frac{p}{m}+\frac{q}{n}-k-\frac{1}{m}\right)} = 0$. This is true, given the two inequalities.

For $y \ge x$, we have

$$\frac{x^p y^q}{(x^m+y^n)^k(x+y)} \le \frac{y^{p+q}}{y^{nk}y} = y^{p+q-nk-1}.$$

Since $p+q > nk+1$, the restriction of the limit on the region converges to 0.

We conclude that the function is differentiable at $(0,0)$ if and only if

$$\frac{p}{m}+\frac{q}{n} > k + \frac{\min\{m,n\}}{mn}, \quad p+q > \min\{m,n\}k+1.$$

Exercise 8.14 (4)

By $f(x,0) = x^{pr-mk}$ and $f(0,y) = y^{qr-nk}$, the function has partial derivatives if and only if $pr-mk > 1$ and $qr-nk > 1$. Moreover, we have $f_x(0,0) = f_y(0,0) = 0$ when the condition is satisfied.

If the function is differentiable at $(0,0)$, then $pr-mk > 1$ and $qr-nk > 1$, the linear approximation is 0, and we have

$$\lim_{x,y\to 0^+} \frac{(x^p+y^q)^r}{(x^m+y^n)^k(x+y)} = 0.$$

Denote

$$\lambda = \min\left\{\frac{p}{m}, \frac{q}{n}\right\}, \quad \mu = \min\{m,n\}, \quad \nu = \min\{p,q\}.$$

As in Exercise 6.23(2), restricting the limit to $x^m = y^n$, we have

$$x^{\lambda mr} \le (x^p+y^q)^r \le 2^r x^{\lambda mr}, \quad (x^m+y^n)^k = 2^k x^{mk}, \quad x^{\frac{\mu}{n}} \le x+y \le 2x^{\frac{\mu}{n}}.$$

Therefore the restriction has limit 0 if and only if $\lim_{x\to 0^+} \dfrac{x^{\lambda mr}}{x^{mk}x^{\frac{\mu}{n}}} = 0$. This means that $\lambda mr > mk + \dfrac{\mu}{n}$, which is the same as

$$r\min\left\{\frac{p}{m}, \frac{q}{n}\right\} > k + \frac{\min\{m,n\}}{mn}.$$

On the other hand, restricting the limit to $x = y$, we have

$$x^{\nu r} \le (x^p + y^q)^r \le 2^r x^{\nu r}, \quad x^{\mu k} \le (x^m + y^n)^k \le 2^k x^{\mu k}, \quad x + y = 2x.$$

Therefore the restriction has limit 0 if and only if $\lim_{x\to 0^+} \dfrac{x^{\nu r}}{x^{\mu k}x} = 0$. This means that $\nu r > \mu k + 1$, which is the same as

$$r\min\{p,q\} > k\min\{m,n\} + 1.$$

Conversely, assume the two inequalities hold. Without loss of generality, assume $m \ge n$. Then the two inequalities mean that

$$r\lambda = r\min\left\{\frac{p}{m}, \frac{q}{n}\right\} > k + \frac{1}{m}, \quad r\nu = r\min\{p,q\} > kn + 1.$$

Then we consider three regions.

For $y \le x^{\frac{m}{n}}$, we have

$$\frac{(x^p + y^q)^r}{(x^m + y^n)^k(x+y)} \le \frac{(x^p + x^{q\frac{m}{n}})^r}{x^{mk}x} \le \frac{2^r x^{\lambda mr}}{x^{mk}x}.$$

Since $\lambda mr - mk - 1 = m\left(r\lambda - k - \dfrac{1}{m}\right) > 0$, the restriction of the limit on the region converges to 0.

For $y \ge x$, we have

$$\frac{(x^p + y^q)^r}{(x^m + y^n)^k(x+y)} \le \frac{(y^p + y^q)^r}{y^{nk}y} \le \frac{2^r y^{\nu r}}{y^{nk}y}.$$

Since $\nu r - nk - 1 > 0$, the restriction of the limit on the region converges to 0.

For $x^{\frac{m}{n}} \le y \le x$, we have

$$y^{nk}x \le (x^m + y^n)^k(x+y) \le 2^k y^{nk} 2x.$$

Moreover, we have

$$\frac{1}{2}(u^r + v^r) \le \max\{u^r, v^r\} \le (u+v)^r \le (2\max\{u,v\})^r = 2^r(\max\{u,v\})^r \le 2^r(u^r + v^r).$$

By substituting $u = x^p$ and $v = y^q$, we see that the restriction of the limit to the region converges to 0 if and only if the limit of $\dfrac{x^{pr} + y^{qr}}{y^{nk}x}$ converges to 0, which is the same as the limits of $\dfrac{x^{pr}}{y^{nk}x}$ and $\dfrac{y^{qr}}{y^{nk}x}$ converge to 0. For fixed x, the maxima and minima of $\dfrac{x^{pr}}{y^{nk}x}$ and $\dfrac{y^{qr}}{y^{nk}x}$ for the y in the range are $x^{pr-nk-1}$, $x^{pr-mk-1}$, $x^{qr-nk-1}$, $x^{qr\frac{m}{n}-mk-1}$. Therefore the limit converges to zero if and only if

$$pr-nk-1>0,\quad pr-mk-1>0,\quad qr-nk-1>0,\quad qr\frac{m}{n}-mk-1>0.$$

Since the two inequalities imply all four inequalities above, we conclude that the two inequalities are necessary and sufficient.

In conclusion, the function is differentiable at $(0,0)$ if and only if

$$r\min\left\{\frac{p}{m},\frac{q}{n}\right\}>k+\frac{\min\{m,n\}}{mn},\quad r\min\{p,q\}>k\min\{m,n\}+1.$$

Exercise 8.15

In Example 8.1.3, we get

$$\|\vec{x}_0+\Delta\vec{x}\|_2^2=\|\vec{x}_0\|_2^2+2\vec{x}_0\cdot\Delta\vec{x}+\|\Delta\vec{x}\|_2^2.$$

So the gradient should give the linear functional $2\vec{x}_0\cdot\Delta\vec{x}$. This means that $\nabla(\vec{x}\cdot\vec{x})=2\vec{x}$.

Exercise 8.16

By the computation in Exercise 8.11, for 2×2 matrix, we have

$$\begin{aligned}\det{}'(A)(H)&=\det(\vec{a}_1,\vec{h}_2)+\det(\vec{h}_1,\vec{a}_2)\\&=a_{11}h_{22}-a_{21}h_{12}+a_{22}h_{11}-a_{12}h_{21}\\&=\begin{pmatrix}a_{22}&-a_{21}\\-a_{12}&a_{11}\end{pmatrix}\cdot\begin{pmatrix}h_{11}&h_{12}\\h_{21}&h_{22}\end{pmatrix}.\end{aligned}$$

Here $\cdot$ is the dot product in $\mathbb{R}^4$. Therefore the gradient

$$\nabla\det=\begin{pmatrix}a_{22}&-a_{21}\\-a_{12}&a_{11}\end{pmatrix}.$$

Exercise 8.17

For $X=\begin{pmatrix}x&y\\z&w\end{pmatrix}$, we have $X^2=\begin{pmatrix}x^2+yz&xy+yw\\xz+zw&yz+w^2\end{pmatrix}$. If we identify the 2×2 matrix X with $(x,y,z,w)^T\in\mathbb{R}^4$, then the Jacobian matrix is

$$(X^2)'=\begin{pmatrix}\dfrac{\partial(x^2+yz)}{\partial x}&\dfrac{\partial(x^2+yz)}{\partial y}&\dfrac{\partial(x^2+yz)}{\partial z}&\dfrac{\partial(x^2+yz)}{\partial w}\\\dfrac{\partial(xy+yw)}{\partial x}&\dfrac{\partial(xy+yw)}{\partial y}&\dfrac{\partial(xy+yw)}{\partial z}&\dfrac{\partial(xy+yw)}{\partial w}\\\dfrac{\partial(xz+zw)}{\partial x}&\dfrac{\partial(xz+zw)}{\partial y}&\dfrac{\partial(xz+zw)}{\partial z}&\dfrac{\partial(xz+zw)}{\partial w}\\\dfrac{\partial(yz+w^2)}{\partial x}&\dfrac{\partial(yz+w^2)}{\partial y}&\dfrac{\partial(yz+w^2)}{\partial z}&\dfrac{\partial(yz+w^2)}{\partial w}\end{pmatrix}=\begin{pmatrix}2x&z&y&0\\y&x+w&0&y\\z&0&x+w&z\\0&z&y&2w\end{pmatrix}.$$

Applying the matrix to the matrix $H = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, which corresponds to the vector $(a, b, c, d)^T \in \mathbb{R}^4$, we get

$$(X^2)'(H) = \begin{pmatrix} 2x & z & y & 0 \\ y & x+w & 0 & y \\ z & 0 & x+w & z \\ 0 & z & y & 2w \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 2xa + zb + yc \\ ya + (x+w)b + yd \\ za + (x+w)c + zd \\ zb + yc + 2wd \end{pmatrix}.$$

Translated back into matrix, this means

$$\begin{aligned} (X^2)'(H) &= \begin{pmatrix} 2xa + zb + yc & ya + (x+w)b + yd \\ za + (x+w)c + zd & zb + yc + 2wd \end{pmatrix} \\ &= \begin{pmatrix} x & y \\ z & w \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x & y \\ z & w \end{pmatrix} = XH + HX. \end{aligned}$$

This recovers the computation in Example 8.1.5.

Exercise 8.18

For $X = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$, we have $X^{-1} = \dfrac{1}{xw - yz} \begin{pmatrix} w & -y \\ -z & x \end{pmatrix}$. If we identify the 2×2 matrix X with $(x, y, z, w)^T \in \mathbb{R}^4$, then the Jacobian matrix is

$$\begin{aligned} (X^{-1})' &= \frac{\partial}{\partial X} \left[\frac{1}{xw - yz} \begin{pmatrix} w \\ -y \\ -z \\ x \end{pmatrix} \right] = \frac{\partial}{\partial X} \begin{pmatrix} w \\ -y \\ -z \\ x \end{pmatrix} + \begin{pmatrix} w \\ -y \\ -z \\ x \end{pmatrix} \frac{\partial}{\partial X} \left[\frac{1}{xw - yz} \right] \\ &= \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} - \frac{1}{(xw - yz)^2} \begin{pmatrix} w \\ -y \\ -z \\ x \end{pmatrix} (w \ -z \ -y \ x) \\ &= \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} - \frac{1}{(xw - yz)^2} \begin{pmatrix} w^2 & -zw & -yw & xw \\ -yw & yz & y^2 & -xy \\ -zw & z^2 & yz & -xz \\ xw & -xz & -xy & x^2 \end{pmatrix} \end{aligned}$$

Then at the identity matrix, we have

$$(X^{-1})'_{\text{at } I} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} - \frac{1}{1^2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = -\text{identity}.$$

This recovers the computation in Exercise 8.13.

Exercise 8.19

Let $\nabla f = (a,b,c)$. Then the condition tells us

$$\begin{aligned}
1 &= \nabla f \cdot \frac{(1,2,2)}{\|(1,2,2)\|} = \frac{1}{3}(a+2b+2c),\\
\sqrt{2} &= \nabla f \cdot \frac{(0,1,-1)}{\|(0,1,-1)\|} = \frac{1}{\sqrt{2}}(b-c),\\
3 &= \nabla f \cdot \frac{(0,0,1)}{\|(0,0,1)\|} = c.
\end{aligned}$$

Solving the system, we get $\nabla f = (-15, 5, 3)$.

Exercise 8.20

Let $\nabla f = a_1\vec{u}_1 + a_2\vec{u}_2 + \cdots + a_n\vec{u}_n$. Then by the orthonormal property, we have $D_{\vec{u}_1} f = \nabla f \cdot \vec{u}_1 = a_1\vec{u}_1\cdot\vec{u}_1 + a_2\vec{u}_2\cdot\vec{u}_1 + \cdots + a_n\vec{u}_n\cdot\vec{u}_1 = a_1 1 + a_2 0 + \cdots + a_n 0 = a_1$. The other coefficients can be similarly derived.

Exercise 8.21 (1)

$$r_x = \frac{x}{\sqrt{x^2+y^2}},\ r_x = \frac{y}{\sqrt{x^2+y^2}},\ \theta_x = \frac{1}{1+\dfrac{y^2}{x^2}}\frac{-y}{x^2} = \frac{-y}{x^2+y^2},\ \theta_y = \frac{1}{1+\dfrac{y^2}{x^2}}\frac{y}{x} = \frac{x}{x^2+y^2}.$$

The Jacobian matrix $\dfrac{\partial(r,\theta)}{\partial(x,y)} = \dfrac{1}{x^2+y^2}\begin{pmatrix} x\sqrt{x^2+y^2} & y\sqrt{x^2+y^2} \\ -y & x \end{pmatrix}$.

The differential $\begin{pmatrix} dr \\ d\theta \end{pmatrix} = \begin{pmatrix} \sqrt{x^2+y^2}(xdx+ydy) \\ -ydx+xdy \end{pmatrix}$.

Exercise 8.21 (2)

The Jacobian matrix $\dfrac{\partial(u_1,u_2,u_3)}{\partial(x_1,x_2,x_3)} = \begin{pmatrix} 1 & 1 & 1 \\ x_2+x_3 & x_3+x_1 & x_1+x_2 \\ x_2x_3 & x_3x_1 & x_1x_2 \end{pmatrix}$.

The differential $\begin{pmatrix} du_1 \\ du_2 \\ du_3 \end{pmatrix} = \begin{pmatrix} dx_1+dx_2+dx_3 \\ (x_2+x_3)dx_1 + (x_3+x_1)dx_2 + (x_1+x_2)dx_3 \\ x_2x_3dx_1 + x_3x_1dx_2 + x_1x_2dx_3 \end{pmatrix}$.

Exercise 8.21 (3)

The Jacobian matrix $\dfrac{\partial(x,y,z)}{\partial(r,\phi,\theta)} = \begin{pmatrix} \sin\phi\cos\theta & r\cos\phi\cos\theta & -r\sin\phi\sin\theta \\ \sin\phi\sin\theta & r\cos\phi\sin\theta & r\sin\phi\cos\theta \\ \cos\phi & -r\sin\phi & 0 \end{pmatrix}$.

The differential $\begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} = \begin{pmatrix} \sin\phi\cos\theta dr + r\cos\phi\cos\theta d\phi - r\sin\phi\sin\theta d\theta \\ \sin\phi\sin\theta dr + r\cos\phi\sin\theta d\phi + r\sin\phi\cos\theta d\theta \\ \cos\phi dr - r\sin\phi d\phi \end{pmatrix}$.

Exercise 8.22

Suppose $\|F(\vec{x})\| < B$. Then $\|\vec{x}\|_2 < \delta$ implies $\|\|\vec{x}\|_2^2 F(\vec{x})\| = \|\vec{x}\|_2^2\|F(\vec{x})\| < \delta B\|\vec{x}\|_2$. This shows that $\|\vec{x}\|_2^2 F(\vec{x})$ is differentiable at $\vec{0}$, with derivatives O.

On the other hand, if we choose $f(\vec{x}) = k$ if k coordinates are rational and $n-k$ coordinates are irrational. Then $\|\vec{x}\|_2^2 f(\vec{x})$ is not continuous along any coordinate except $\vec{0}$. Thus the function has no partial derivatives away from $\vec{0}$.

Exercise 8.23

The function $f(x,y)=\begin{cases}(x^2+y^2)\sin\dfrac{1}{x^2+y^2} & \text{if}(x,y)\neq(0,0)\\ 0 & \text{if}(x,y)=(0,0)\end{cases}$ in Exercise 8.14 is differentiable everywhere. But $f(x,0)=x^2\sin\frac{1}{x^2}$, and $f_x(x,0)=2x\sin\frac{1}{x^2}-\frac{2}{x}\cos\frac{1}{x^2}$ is not continuous at $x=0$. The other partial derivative is also not continuous.

Exercise 8.24

By the existence of $f_x(x_0,y_0)$, for any $\epsilon>0$, there is $\delta>0$, such that $|\Delta x|<\delta$ implies

$$|f(x,y_0)-f(x_0,y_0)-f_x(x_0,y_0)\Delta x|\leq\epsilon|\Delta x|.$$

By mean value theorem, we have $f(x,y)-f(x,y_0)=f_y(x,d)\Delta y$. Then by the continuity of f_y at (x_0,y_0), there is $\delta'>0$, such that $|\Delta x|<\delta'$, $|\Delta y|<\delta'$ implies $|f_y(x,d)-f_y(x_0,y_0)|<\epsilon$. Therefore

$$|f(x,y)-f(x,y_0)-f_y(x_0,y_0)\Delta y|=|(f_y(x,d)-f_y(x_0,y_0))\Delta y|\leq\epsilon|\Delta y|.$$

Combining the two estimations, for $|\Delta x|<\min\{\delta,\delta'\}$, $|\Delta y|<\min\{\delta,\delta'\}$, we have

$$\begin{aligned}&|f(x,y)-f(x_0,y_0)-f_x(x_0,y_0)\Delta x-f_y(x_0,y_0)\Delta y|\\ \leq&|f(x,y_0)-f(x_0,y_0)-f_x(x_0,y_0)\Delta x|+|f(x,y)-f(x,y_0)-f_y(x_0,y_0)\Delta y|\\ \leq&\epsilon|\Delta x|+\epsilon|\Delta y|.\end{aligned}$$

This proves the differentiability of f at (x_0,y_0).

In general, if one partial derivative exist and the other partial derivatives are continuous, then the function is differentiable.

Exercise 8.25

If $F(x,y)$ is differentiable, then the composition (or restriction) $F(x,x)=f'(x)$ is differentiable, which means the existence of second order derivative.

Conversely, suppose f has second order derivative. By Proposition 3.4.2 we have

$$f(x)=f(x_0)+f'(x_0)\Delta x+\frac{1}{2}f''(x_0)\Delta x^2+R(x),\quad \lim_{x\to x_0}\frac{R(x)}{\Delta x^2}=0.$$

We also note that $R(x)$ has second order derivative, and $R(x_0)=R'(x_0)=R''(x_0)=0$. The Taylor expansion implies

$$\begin{aligned}f(x)-f(y)&=f'(x_0)(\Delta x-\Delta y)+\frac{1}{2}f''(x_0)(\Delta x^2-\Delta y^2)+R(x)-R(y)\\ &=f'(x_0)(x-y)+\frac{1}{2}f''(x_0)(x-y)(\Delta x+\Delta y)+R(x)-R(y).\end{aligned}$$

Then for $x\neq y$ and (x,y) near (x_0,x_0), we get

$$\frac{f(x)-f(y)}{x-y}=f'(x_0)+\frac{1}{2}f''(x_0)(\Delta x+\Delta y)+\frac{R(x)-R(y)}{x-y}.$$

If we can show that $\dfrac{R(x)-R(y)}{x-y}=o(\|(\Delta x,\Delta y)\|)$, then we find $F(x,y)$ is differentiable at (x_0,x_0) with

$$F'(x_0,x_0)(u,v)=\frac{1}{2}f''(x_0)(u+v).$$

By the mean value theorem, we have $\dfrac{R(x)-R(y)}{x-y}=R'(c)$, where c is between x and y. Using $R'(x_0)=R''(x_0)=0$, we further have

$$R'(c)=R'(x_0)+R''(x_0)(c-x_0)+o(c-x_0)=o(c-x_0).$$

Since c is between x and y, we have $|c-x_0|\le\max\{|x-x_0|,|y-x_0|\}=\|(\Delta x,\Delta y)\|_\infty$. Therefore

$$\frac{R(x)-R(y)}{x-y}=R'(c)=o(\|(\Delta x,\Delta y)\|_\infty).$$

This completes the proof that F is differentiable at (x_0,x_0).

In case $x_0\ne y_0$, both $f(x)-f(y)$ and $x-y$ are differentiable at (x_0,y_0) and $x-y\ne 0$ near (x_0,y_0). Therefore the quotient $F(x,y)$ is differentiable at (x_0,y_0).

Additional: F' is continuous if and only if f'' is continuous.

We already know that, if f has second order derivative, then

$$F_x(x,y)=\begin{cases}\dfrac{f'(x)(x-y)-f(x)+f(y)}{(x-y)^2}, & \text{if } x\ne y,\\[2mm] \dfrac{1}{2}f''(x), & \text{if } x=y,\end{cases}$$

and we have the similar formula for F_y. The continuity of F' means the continuity of F_x and F_y. The formula tells us that the continuity of $F_x(x,x)$ already means f'' is continuous.

Conversely, suppose $f''(x)$ is continuous. We want to show that F_x is continuous. The continuity of f' already implies the continuity of F_x at (x_0,y_0) whenever $x_0\ne y_0$. It remains to show the continuity at (x_0,x_0).

For $x\ne y$, we have

$$\begin{aligned}F_x(x,y)-\frac{1}{2}f''(x_0)&=\frac{f(y)-f(x)-f'(x)(y-x)-\frac{1}{2}f''(x_0)(y-x)^2}{(y-x)^2}\\&=\frac{r_1(y,x)}{(y-x)^2}=\frac{r_1(y,x)-r_1(x,x)}{(y-x)^2-(x-x)^2}\\&=\frac{D_1r_1(c_1,x)}{2(c_1-x)}=\frac{f'(c_1)-f'(x)-f''(x_0)(c_1-x)}{2(c-x)}\\&=\frac{r_2(c_1,x)}{2(c-x)}=\frac{r_2(c_1,x)-r_2(x,x)}{2(c-x)}\\&=\frac{D_1r_2(c_2,x)}{2}=f''(c_2)-f''(x_0).\end{aligned}$$

Since c_1 is between y and x_0, and c_2 is between c_1 and x_0, the continuity of f'' shows that the limit above as $(x,y)\to(x_0,y_0)$ is 0.

Exercise 8.31

We may copy the argument for Exercise 8.8 word by word. The argument using small o notation is equivalent to using Exercises 8.27 and 8.29.

Let

$$F(\vec{x}) = F(\vec{x}_0) + F'(\vec{x}_0)(\Delta\vec{x}) + R_1, \qquad R_1 = o(\|\Delta\vec{x}\|);$$
$$G(\vec{x}) = G(\vec{x}_0) + G'(\vec{x}_0)(\Delta\vec{x}) + R_2, \qquad R_2 = o(\|\Delta\vec{x}\|).$$

Then

$$\begin{aligned}
B(F(\vec{x}), G(\vec{x})) &= B(F(\vec{x}_0) + F'(\vec{x}_0)(\Delta\vec{x}) + R_1, G(\vec{x}_0) + G'(\vec{x}_0)(\Delta\vec{x}) + R_2) \\
&= B(F(\vec{x}_0), G(\vec{x}_0)) + B(F'(\vec{x}_0)(\Delta\vec{x}), G(\vec{x}_0)) + B(F(\vec{x}_0), G'(\vec{x}_0)(\Delta\vec{x})) + R, \\
R &= B(F(\vec{x}_0), R_2) + B(R_1, G(\vec{x}_0)) + B(F'(\vec{x}_0)(\Delta\vec{x}), R_2) + B(R_1, G'(\vec{x}_0)(\Delta\vec{x})) \\
&\quad + B(F'(\vec{x}_0)(\Delta\vec{x}), G'(\vec{x}_0)(\Delta\vec{x})) + B(R_1, R_2).
\end{aligned}$$

By $\|B(\vec{u}, \vec{v})\| \le \|B\|\|\vec{u}\|\|\vec{v}\|$, we have

$$\begin{aligned}
\|B(F(\vec{x}_0), R_2)\| &\le \|B\|\|F(\vec{x}_0)\|\|R_2\|, \\
\|B(R_1, G(\vec{x}_0))\| &\le \|B\|\|G(\vec{x}_0)\|\|R_1\|, \\
\|B(F'(\vec{x}_0)(\Delta\vec{x}), R_2)\| &\le \|B\|\|F'(\vec{x}_0)(\Delta\vec{x})\|\|R_2\| \le \|B\|\|F'(\vec{x}_0)\|\|\Delta\vec{x}\|\|R_2\|, \\
\|B(R_1, G'(\vec{x}_0)(\Delta\vec{x}))\| &\le \|B\|\|G'(\vec{x}_0)(\Delta\vec{x})\|\|R_1\| \le \|B\|\|G'(\vec{x}_0)\|\|\Delta\vec{x}\|\|R_1\|, \\
\|B(F'(\vec{x}_0)(\Delta\vec{x}), G'(\vec{x}_0)(\Delta\vec{x}))\| &\le \|B\|\|F'(\vec{x}_0)(\Delta\vec{x})\|\|G'(\vec{x}_0)(\Delta\vec{x})\| \le \|B\|\|F'(\vec{x}_0)\|\|G'(\vec{x}_0)\|\|\Delta\vec{x}\|^2, \\
\|B(R_1, R_2)\| &\le \|B\|\|R_1\|\|R_2\|.
\end{aligned}$$

Then by $R_1 = o(\Delta\vec{x})$ and $R_2 = o(\Delta\vec{x})$, the estimations above imply that $B(F(\vec{x}_0), R_2) = o(\|\Delta\vec{x}\|)$, $B(F'(\vec{x}_0), R_2) = o(\|\Delta\vec{x}\|)$, and the other four terms in R are $o(\|\Delta\vec{x}\|^2)$. Therefore $R = o(\|\Delta\vec{x}\|)$, and $B(F(\vec{x}_0), G(\vec{x}_0)) + B(F'(\vec{x}_0)(\Delta\vec{x}), G(\vec{x}_0)) + B(F(\vec{x}_0), G'(\vec{x}_0)(\Delta\vec{x}))$ is the linear approximation of $B(F(\vec{x}), G(\vec{x}))$ at $\vec{x}_0$. In particular, we get

$$B(F, G)'(\vec{x}_0)(\vec{v}) = B(F'(\vec{x}_0)(\vec{v}), G(\vec{x}_0)) + B(F(\vec{x}_0), G'(\vec{x}_0)(\vec{v})).$$

[Alternative: Argue as special case of Exercise 8.28]

Let $\vec{a} = F(\vec{x}_0)$, $\vec{b} = G(\vec{x}_0)$, $L = F'(\vec{x}_0)$ and $K = G'(\vec{x}_0)$. Then

$$F(\vec{x}) \sim_{\|\Delta\vec{x}\|} \vec{a} + L(\Delta\vec{x}), \quad G(\vec{x}) \sim_{\|\Delta\vec{x}\|} \vec{b} + K(\Delta\vec{x}).$$

By Exercise 8.28, we have

$$\begin{aligned}
B(F(\vec{x}), G(\vec{x})) &\sim_{\|\Delta\vec{x}\|} B(\vec{a} + L(\Delta\vec{x}), \vec{b} + K(\Delta\vec{x})) \\
&= B(\vec{a}, \vec{b}) + B(L(\Delta\vec{x}), \vec{b}) + B(\vec{a}, K(\Delta\vec{x})) + B(L(\Delta\vec{x}), K(\Delta\vec{x})).
\end{aligned}$$

Since

$$\|B(L(\Delta\vec{x}), K(\Delta\vec{x}))\| \le \|B\|\|L(\Delta\vec{x})\|\|K(\Delta\vec{x}))\| \le \|B\|\|L\|\|K\|\|\Delta\vec{x})\|^2,$$

we get $B(L(\Delta\vec{x}),K(\Delta\vec{x}))\sim_{\|\Delta\vec{x}\|}0$. Then by Exercise 8.27, we get

$$\begin{aligned}&B(\vec{a},\vec{b})+B(L(\Delta\vec{x}),\vec{b})+B(\vec{a},K(\Delta\vec{x}))+B(L(\Delta\vec{x}),K(\Delta\vec{x}))\\ \sim_{\|\Delta\vec{x}\|}\ &B(\vec{a},\vec{b})+B(L(\Delta\vec{x}),\vec{b})+B(\vec{a},K(\Delta\vec{x})).\end{aligned}$$

Further by Exercise 8.29, we get

$$B(F(\vec{x}),G(\vec{x}))\sim_{\|\Delta\vec{x}\|}B(\vec{a},\vec{b})+B(L(\Delta\vec{x}),\vec{b})+B(\vec{a},K(\Delta\vec{x})).$$

Then the linear part

$$B(L(\Delta\vec{x}),\vec{b})+B(\vec{a},K(\Delta\vec{x}))=B(F'(\vec{x}_0)(\Delta\vec{x}),G(\vec{x}_0))+B(F(\vec{x}_0),G'(\vec{x}_0)(\Delta\vec{x}))$$

of the right side is then the derivative $B(F,G)'(\vec{x}_0)(\Delta\vec{x})$.

Exercise 8.32

If G is a multilinear map, and $F_1,\dots,F_k$ are differentiable maps, then

$$G(F_1,F_2,\dots,F_k)'=G(F_1',F_2,\dots,F_k)+G(F_1,F_2',\dots,F_k)+\cdots+G(F_1,F_2,\dots,F_k').$$

Exercise 8.33 (1)

We have

$$F(\vec{x})=F(\vec{x}_0)+F'(\vec{x}_0)(\Delta\vec{x})+R_1(\vec{x})=F'(\vec{x}_0)(\Delta\vec{x})+R_1(\vec{x}),\quad R_1(\vec{x})=o(\|\Delta\vec{x}\|),$$

and similarly

$$G(\vec{x})=G(\vec{x}_0)+G'(\vec{x}_0)(\Delta\vec{x})+R_2(\vec{x})=G'(\vec{x}_0)(\Delta\vec{x})+R_2(\vec{x}),\quad R_2(\vec{x})=o(\|\Delta\vec{x}\|).$$

Then

$$\begin{aligned}B(F(\vec{x}),G(\vec{x}))=&\,B(F'(\vec{x}_0)(\Delta\vec{x}),G'(\vec{x}_0)(\Delta\vec{x}))\\&+B(F'(\vec{x}_0)(\Delta\vec{x}),R_2(\vec{x}))+B(R_1(\vec{x}),G'(\vec{x}_0)(\Delta\vec{x}))+B(R_1(\vec{x}),R_2(\vec{x})).\end{aligned}$$

For any $\epsilon>0$, there is $\delta>0$, such that

$$\|\Delta\vec{x}\|<\delta\implies\|R_1(\vec{x})\|<\epsilon\|\Delta\vec{x}\|,\ \|R_2(\vec{x})\|<\epsilon\|\Delta\vec{x}\|.$$

Then

$$\begin{aligned}\|B(F'(\vec{x}_0)(\Delta\vec{x}),R_2(\vec{x}))\|&\le\|B\|\|F'(\vec{x}_0)(\Delta\vec{x})\|\|R_2(\vec{x})\|\le\|B\|\|F'(\vec{x}_0)\|\|\Delta\vec{x}\|\epsilon\|\Delta\vec{x}\|,\\ \|B(R_1(\vec{x}),G'(\vec{x}_0)(\Delta\vec{x}))\|&\le\|B\|\|R_1(\vec{x})\|\|G'(\vec{x}_0)(\Delta\vec{x})\|\le\|B\|\epsilon\|\Delta\vec{x}\|\|G'(\vec{x}_0)\|\|\Delta\vec{x}\|,\\ \|B(R_1(\vec{x}),R_2(\vec{x}))\|&\le\|B\|\|R_1(\vec{x})\|\|R_2(\vec{x})\|\le\|B\|\epsilon\|\Delta\vec{x}\|\epsilon\|\Delta\vec{x}\|.\end{aligned}$$

This implies

$$\|B(F(\vec{x}),G(\vec{x}))-B(F'(\vec{x}_0)(\Delta\vec{x}),G'(\vec{x}_0)(\Delta\vec{x}))\|\le\epsilon\|B\|(\|F'(\vec{x}_0)\|+\|G'(\vec{x}_0)\|+\epsilon)\|\Delta\vec{x}\|,$$

and proves

$$B(F,G) \sim_{\|\Delta\vec{x}\|^2} B(F'(\vec{x}_0)(\Delta x), G'(\vec{x}_0)(\Delta x)).$$

In general, if $G(\vec{y}_1,\dots,\vec{y}_k)$ is multilinear, and $F_i(\vec{x}_0)=\vec{0}$, then

$$G(F_1,\dots,F_k) \sim_{\|\Delta\vec{x}\|^k} G(F_1'(\vec{x}_0)(\Delta x),\dots,F_k'(\vec{x}_0)(\Delta x)).$$

Exercise 8.33 (2)

We have $Q(\vec{y})=B(\vec{y},\vec{y})$ for a symmetric bilinear map B. Then by the first part, we have

$$Q(F) \sim_{\|\Delta\vec{x}\|^2} B(F'(\vec{x}_0)(\Delta x), F'(\vec{x}_0)(\Delta x)) = Q(F'(\vec{x}_0)(\Delta x)).$$

In general, if $G(\vec{y})$ is k-th order, and $F_i(\vec{x}_0)=\vec{0}$, then

$$G(F(\vec{x})) \sim_{\|\Delta\vec{x}\|^k} G(F'(\vec{x}_0)(\Delta x)).$$

Exercise 8.34

The gradient is defined by $f'(\vec{x}_0)(\vec{v})=\nabla f(\vec{x}_0)\cdot\vec{v}$ for all $\vec{v}$. The calculation of the gradient follows from the calculation of the derivative.

By $(f+g)'(\vec{x}_0)=f'(\vec{x}_0)+g'(\vec{x}_0)$, we get

$$\nabla(f+g)(\vec{x}_0)\cdot\vec{v}=\nabla f(\vec{x}_0)\cdot\vec{v}+\nabla g(\vec{x}_0)\cdot\vec{v}=(\nabla f(\vec{x}_0)+\nabla g(\vec{x}_0))\cdot\vec{v}.$$

Since this holds for all $\vec{v}$, we get $\nabla(f+g)=\nabla f+\nabla g$.

By $(fg)'(\vec{x}_0)(\vec{v})=g(\vec{x}_0)f'(\vec{x}_0)(\vec{v})+f(\vec{x}_0)g'(\vec{x}_0)(\vec{v})$, we get

$$\nabla(fg)(\vec{x}_0)\cdot\vec{v}=g(\vec{x}_0)(\nabla f(\vec{x}_0)\cdot\vec{v})+f(\vec{x}_0)(\nabla g(\vec{x}_0)\cdot\vec{v})=(g(\vec{x}_0)\nabla f(\vec{x}_0)+f(\vec{x}_0)\nabla g(\vec{x}_0))\cdot\vec{v}.$$

Since this holds for all $\vec{v}$, we get $\nabla(fg)=g\nabla f+f\nabla g$.

We have

$$\begin{aligned}(F\cdot G)'(\vec{x}_0)(\vec{v}) &= F'(\vec{x}_0)(\vec{v})\cdot G(\vec{x}_0)+F(\vec{x}_0)\cdot G'(\vec{x}_0)(\vec{v})\\ &= \vec{v}\cdot F'(\vec{x}_0)^*(G(\vec{x}_0))+G'(\vec{x}_0)^*(F(\vec{x}_0))\cdot\vec{v}\\ &= [F'(\vec{x}_0)^*(G(\vec{x}_0))+G'(\vec{x}_0)^*(F(\vec{x}_0))]\cdot\vec{v}.\end{aligned}$$

Since this holds for all $\vec{v}$, we get

$$\nabla(F\cdot G)(\vec{x}_0)=F'(\vec{x}_0)^*(G(\vec{x}_0))+G'(\vec{x}_0)^*(F(\vec{x}_0)),$$

or

$$\nabla(F\cdot G)=F'^*(G)+G'^*(F).$$

Exercise 8.35

We may prove the chain rule similar to the single variable case. This means applying Exercise 8.30 to the case P is the linear approximation of F, with $u=\|\Delta\vec{x}\|$, and Q is the linear approximation of G, with $v=\|\Delta\vec{y}\|$.

The following is a more direct proof.

Let $\vec{y}_0 = F(\vec{x}_0)$, $\vec{z}_0 = G(\vec{y}_0)$, $L = F'(\vec{x}_0)$, $K = G'(\vec{y}_0)$. For any $\epsilon > 0$, there is $\mu > 0$, such that $\|\Delta\vec{y}\| < \mu$ implies

$$\|G(\vec{y}) - \vec{z}_0 - K(\Delta\vec{y})\| \le \epsilon\|\Delta\vec{y}\|.$$

Then there is $\delta > 0$, such that $\|\Delta\vec{x}\| < \delta$ implies

$$\|F(\vec{x}) - \vec{y}_0 - L(\Delta\vec{x})\| \le \epsilon\|\Delta\vec{x}\|.$$

Note that this implies $\|F(\vec{x}) - \vec{y}_0\| \le (\|L\| + \epsilon)\|\Delta\vec{x}\| < (\|L\| + \epsilon)\delta$. By choosing $\delta < \dfrac{\mu}{\|L\| + \epsilon}$ to begin with, we also know $\|F(\vec{x}) - \vec{y}_0\| < \mu$, so that

$$\|G(F(\vec{x})) - \vec{z}_0 - K(F(\vec{x}) - \vec{y}_0)\| \le \epsilon\|F(\vec{x}) - \vec{y}_0\| \le \epsilon(\|L\| + \epsilon)\|\Delta\vec{x}\|.$$

Then

$$\begin{aligned}
&\|G(F(\vec{x})) - \vec{z}_0 - K(L(\Delta\vec{x}))\| \\
\le&\|G(F(\vec{x})) - \vec{z}_0 - K(F(\vec{x}) - \vec{y}_0)\| + \|K(F(\vec{x}) - \vec{y}_0 - L(\Delta\vec{x}))\| \\
\le&\epsilon(\|L\| + \epsilon)\|\Delta\vec{x}\| + \|K\|\|F(\vec{x}) - \vec{y}_0 - L(\Delta\vec{x})\| \le \epsilon(\|L\| + \epsilon + \|K\|)\|\Delta\vec{x}\|.
\end{aligned}$$

This implies that $\vec{z}_0 + K(L(\Delta\vec{x}))$ is a linear approximation of $G(F(\vec{x}))$ at $\vec{x}_0$ and $(G \circ F)'(\vec{x}_0) = K \circ L$.

Exercise 8.36

By restricting (chain rule used here) to $\vec{x} = t\vec{v}$, we see that $f(\|\vec{x}\|_2)$ is differentiable away from $\vec{0}$ if and only if $f(t)$ is differentiable for $t > 0$.

Now consider the differentiability of $f(\|\vec{x}\|_2)$ at $\vec{0}$. First, since differentiable functions are continuous, we must have $f(0) = \lim_{t\to 0^+} f(t)$. Suppose the linear functional $l(\vec{v})$ is the derivative of $f(\|\vec{x}\|_2)$ at $\vec{0}$. Then for any $\epsilon > 0$, there is $\delta > 0$, such that $\|\vec{x}\|_2 < \delta$ implies $|f(\|\vec{x}\|_2) - f(0) - l(\vec{x})| \le \epsilon\|\vec{x}\|_2$. By applying the condition to $-\vec{x}$, we find $\|\vec{x}\|_2 < \delta$ implies $|f(\|\vec{x}\|_2) - f(0) + l(\vec{x})| \le \epsilon\|\vec{x}\|_2$. Thus $\|\vec{x}\|_2 < \delta$ implies $|2l(\vec{x})| \le |f(\|\vec{x}\|_2) - f(0) - l(\vec{x})| + |f(\|\vec{x}\|_2) - f(0) + l(\vec{x})| \le 2\epsilon\|\vec{x}\|_2$. Thus $|l(\vec{x})| \le \epsilon\|\vec{x}\|_2$ for any $\vec{x}$ satisfying $\|\vec{x}\|_2 < \delta$. This further implies $|l(c\vec{x})| \le \epsilon\|c\vec{x}\|_2$, where $c\vec{x}$ can be any vector of arbitrary length. Therefore we must have $l(c\vec{x}) = 0$ for all $\vec{x}$.

Thus the differentiability at $\vec{0}$ means that for any $\epsilon > 0$, there is $\delta > 0$, such that $\|\vec{x}\|_2 < \delta$ implies $|f(\|\vec{x}\|_2) - f(0)| \le \epsilon\|\vec{x}\|_2$. In other words, $0 < t < \delta$ implies $|f(t) - f(0)| \le \epsilon t$. This means $\lim_{t\to 0^+} \dfrac{f(t) - f(0)}{t} = 0$. Thus we conclude that $f(\|\vec{x}\|_2)$ is differentiable at $\vec{0}$ if and only if $f'_+(0) = 0$.

Exercise 8.37

By the chain rule, $F \circ G = id$ and $G \circ F = id$ implies $F' \circ G' = I$ and $G' \circ F' = I$. Thus the linear transforms F' and G' are inverse to each other.

Exercise 8.38

By taking the derivative of $f(t,t^2)=1$, we get $f_x(t,t^2)+f_y(t,t^2)2t=0$. Then by $f_x(t,t^2)=t$, we get $f_y(t,t^2)=-\dfrac{1}{2}$.

Exercise 8.39

Let $\nabla f=(a,b)$. Then $\phi'=(1,2)$ and $\psi'=(2,1)$ at $(1,1)$, and the condition tells us

$$2=\nabla f\cdot(1,2)=a+2b,\quad 3=\nabla f\cdot(2,1)=2a+b.$$

Solving the system, we get $\nabla f=\dfrac{1}{3}(4,1)$.

Exercise 8.40

We have $f_r=f_x\cos\theta+f_y\sin\theta$ and $f_\theta=-f_xr\sin\theta+f_y\cos\theta$. Then $D_{\vec{e}_r}f=\nabla f\cdot\vec{e}_r=f_x\cos\theta+f_y\sin\theta=f_r$ and $D_{\vec{e}_\theta}f=\nabla f\cdot\vec{e}_\theta=-f_x\sin\theta+f_y\cos\theta=r^{-1}f_\theta$. By Exercise 6.1.32, we have $\nabla f=f_r\vec{e}_r+r^{-1}f_\theta\vec{e}_\theta$.

Exercise 8.41

We have $f_r=f_x\cos\theta+f_y\sin\theta=r^{-1}(xf_x+yf_y)$. Therefore f is independent of $r\iff f_r=0\iff xf_x+yf_y=0$.

We also have $f_\theta=-f_xr\sin\theta+f_yr\cos\theta=-yf_x+xf_y$. Therefore f is independent of $\theta\iff f_\theta=0\iff -yf_x+xf_y=0$.

Exercise 8.42

$$\begin{aligned}xg_x+yg_y+zg_z=&x\left(f_v\frac{\sqrt{z}}{2\sqrt{x}}+f_w\frac{\sqrt{y}}{2\sqrt{x}}\right)+y\left(f_w\frac{\sqrt{x}}{2\sqrt{y}}+f_u\frac{\sqrt{z}}{2\sqrt{y}}\right)+z\left(f_u\frac{\sqrt{y}}{2\sqrt{z}}+f_v\frac{\sqrt{x}}{2\sqrt{z}}\right)\\=&\left(y\frac{\sqrt{z}}{2\sqrt{y}}+z\frac{\sqrt{y}}{2\sqrt{z}}\right)f_u+\left(z\frac{\sqrt{x}}{2\sqrt{z}}+x\frac{\sqrt{z}}{2\sqrt{x}}\right)f_v+\left(x\frac{\sqrt{y}}{2\sqrt{x}}+y\frac{\sqrt{x}}{2\sqrt{y}}\right)f_w\\=&\sqrt{yz}f_u+\sqrt{zx}f_v+\sqrt{xy}f_w=uf_u+vf_v+wf_w.\end{aligned}$$

Exercise 8.43

If $f(x,y)=\phi(xy)$, then $f_x=y\phi'(xy)$, $f_y=x\phi'(xy)$. Therefore $xf_x=yf_y$.

If $f(x,y)=\phi\left(\dfrac{y}{x}\right)$, then $f_x=-\dfrac{y}{x^2}\phi'\left(\dfrac{y}{x}\right)$, $f_y=\dfrac{1}{x}\phi'\left(\dfrac{y}{x}\right)$. Therefore $xf_x+yf_y=0$.

Exercise 8.44

We have

$$(g\circ f)'(\vec{x}_0)(\vec{v})=g'(f(\vec{x}_0))(f'(\vec{x}_0)(\vec{v}))=g'(f(\vec{x}_0))(\nabla f(\vec{x}_0)\cdot\vec{v})=[g'(f(\vec{x}_0))\nabla f(\vec{x}_0)]\cdot\vec{v}.$$

Here the last equality is due to the following reason: Both $g'(f(\vec{x}_0))$ and $\nabla f(\vec{x}_0)\cdot\vec{v}$ are actually numbers. Therefore the application of the former (as linear transform) to the later (as vector) is simply multiplication of two numbers. Since the equality holds for all $\vec{v}$, we get

$$\nabla(g\circ f)(\vec{x}_0)=g'(f(\vec{x}_0))\nabla f(\vec{x}_0),\ \text{or}\ \nabla(g\circ f)=(g'\circ f)\nabla f.$$

For the more general $g(F(\vec{x}))$, we have

$$(g\circ F)'(\vec{x}_0)(\vec{v})=g'(F(\vec{x}_0))(F'(\vec{x}_0)(\vec{v}))=\nabla g(F(\vec{x}_0))\cdot F'(\vec{x}_0)(\vec{v})=F'(\vec{x}_0)^*(\nabla g(F(\vec{x}_0)))\cdot\vec{v}.$$

Since the equality holds for all $\vec{v}$, we get

$$\nabla(g \circ F)(\vec{x}_0) = F'(\vec{x}_0)^*(\nabla g(F(\vec{x}_0))), \text{ or } \nabla(g \circ F) = F'^*(\nabla g \circ F).$$

Exercise 8.45

The derivatives of $X \to A^{-1}X$, $X \to XA^{-1}$ are $H \to A^{-1}H$, $H \to HA^{-1}$ because the two maps are linear. By Exercise 8.12, the derivative of $X \to X^{-1}$ at I is $H \to -H$. Thus the composition $X \to (A^{-1}X)^{-1}A^{-1} = X^{-1}$ has the derivative $(-A^{-1}H)A^{-1} = -A^{-1}HA^{-1}$ at A.

Exercise 8.46

The function is $f(x, y) = \dfrac{x^2 y}{x^2 + y^2}$ for $(x, y) \neq (0, 0)$ and $f(0, 0) = 0$. Since $\nabla f(0, 0) = (0, 0)$, we always have $\nabla f(0, 0) \cdot \vec{v} = 0$. On the other hand, the direct calculation shows that $D_{\vec{v}} f = a^2 b \neq 0$ is not zero for most $\vec{v} = (a, b)$. So the function fails the formula $D_{\vec{v}} f = \nabla f \cdot \vec{v}$.

Since the formula $D_{\vec{v}} f = \nabla f \cdot \vec{v}$ is a special case of the chain rule, the failure of the formula $D_{\vec{v}} f = \nabla f \cdot \vec{v}$ implies the failure of the chain rule. Since the straight line is always differentiable, the failure must be due to the non-differentiability of f.

Exercise 8.47

As explained in Example 8.1.13, the function in Example 6.2.3 satisfies the equality $D_{\vec{v}} f(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot \vec{v}$ for all $\vec{v}$. As explained in Exercise 8.46, this means that $f(\phi(t))$ satisfies the chain rule for all straight lines $\phi(t)$. Yet the function is not differentiable.

Exercise 8.48

The chain rule formula is a formula for the partial derivatives. The partial derivative is taken for a specific variable, while keeping all the other variables constant. Therefore in the chain rule formula for $F \circ G$, only the differentiability of G along one particular variable is used. Since this differentiability is equivalent to the existence of the corresponding partial derivative, we do not expect a counterexample where G only has partial derivatives but is not differentiable.

Exercise 8.49

By the equivalence of norms, we will conclude $\|F(\vec{b}) - F(\vec{a})\| \leq C\|F'(\vec{c})\|\|\vec{b} - \vec{a}\|$ for some $C > 0$ in Proposition 8.2.2, where C is independent of the functions and variables (it depends only on the norms of the Euclidean spaces).

Exercise 8.50

If $f_x(x, y) = 0$ on an open subset U, such that the intersection of U with any x-line $y = c$ is an interval, then $f(x, y) = g(y)$ for a function g of y.

More generally, suppose $f_{\vec{x}}(\vec{x}, \vec{y})$ on an open subset U, such that the intersection of U with any plane $\vec{y} = \vec{c}$ is path connected, then $f(\vec{x}, \vec{y}) = g(\vec{y})$ for a function g of $\vec{y}$.