and set $A = kC + 1$. Then if $x \leq -A$ we have

$$\begin{aligned} p(x) &= x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0 \\ &\leq (-A)^k + CA^{k-1} + \cdots + C \leq -A^k + kCA^{k-1} \\ &= A^{k-1}(kC - A) = -A^{k-1} \leq 0. \end{aligned}$$

Similarly, if $x \geq A$ we have

$$\begin{aligned} p(x) &= x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0 \\ &\geq (A)^k - CA^{k-1} - \cdots - C \geq A^k - kCA^{k-1} \\ &= A^{k-1}(A - kC) = A^{k-1} \geq 0. \end{aligned}$$

Since $p : [-A, A] \to \mathbb{R}$ is a continuous function (corollary 1.5.30) and we have $p(-A) \leq 0$ and $p(A) \geq 0$, then by the intermediate value theorem there exists $x_0 \in [-A, A]$ such that $p(x_0) = 0$.

**1.7.1** a. $f(a) = 0$, $f'(a) = \cos(a) = 1$ so the tangent is $g(x) = x$.

b. $f(a) = \frac{1}{2}$, $f'(a) = -\sin(a) = -\frac{\sqrt{3}}{2}$ so the tangent is

$$g(x) = -\frac{\sqrt{3}}{2}(x - \frac{\pi}{3}) + \frac{1}{2}.$$

c. $f(a) = 1$, $f'(a) = -\sin(a) = 0$ so the tangent is $g(x) = 1$.

d. $f(a) = 2$, $f'(a) = -\frac{1}{a^2} = -4$ so the tangent is

$$g(x) = -4(x - 1/2) + 2 = -4x + 4.$$

**1.7.2** We need to find $a$ such that if the graph of $g$ is the tangent at $a$, then $g(0) = 0$. Since the tangent is

$$g(x) = e^{-a} - e^{-a}(x - a),$$

we have

$$g(0) = e^{-a} + ae^{-a} = 0,$$

so

$$e^{-a}(1 + a) = 0, \quad \text{which gives} \quad a = -1.$$

**1.7.3** a. $f'(x) = \Big(3\sin^2(x^2 + \cos x)\Big)\Big(\cos(x^2 + \cos x)\Big)\Big(2x - \sin x\Big)$

b. $f'(x) = \Big(2\cos((x + \sin x)^2)\Big)\Big(-\sin((x + \sin x)^2)\Big)\Big(2(x + \sin x)\Big)\Big(1 + \cos x\Big)$

c. $f'(x) = \Big((\cos x)^5 + \sin x\Big)\Big(4(\cos x)^3\Big)(-\sin(x)) = (\cos x)^5 - 4(\sin x)^2(\cos x)^3$

d. $f'(x) = 3(x + \sin^4 x)^2(1 + 4\sin^3 x \cos x)$

e. $f'(x) = \dfrac{\sin^3 x(\cos x^2 * 2x)}{2 + \sin(x)} + \dfrac{\sin x^2(3\sin^2 x \cos x)}{2 + \sin(x)} - \dfrac{(\sin x^2 \sin^3 x)(\cos x)}{\big(2 + \sin(x)\big)^2}$

f. $f'(x) = \cos\left(\dfrac{x^3}{\sin x^2}\right)\left(\dfrac{3x^2}{\sin x^2} - \dfrac{(x^3)(\cos x^2 * 2x)}{(\sin x^2)^2}\right)$

**1.7.4** a. If $f(x) = |x|^{3/2}$, then

$$f'(0) = \lim_{h\to 0} \frac{|h|^{3/2}}{h} = \lim_{h\to 0} |h|^{1/2} = 0,$$

so the derivative does exist. But

$$f(0+h) - f(0) - hf'(0) = |h|^{3/2}$$

is larger than $h^2$, since the limit

$$\lim_{h\to 0} \frac{|h|^{3/2}}{h^2} = \lim_{h\to 0} |h|^{-1/2}$$

is infinite.

b. If $f(x) = x \ln |x|$, then the limit

$$f'(0) = \lim_{h\to 0} \frac{h \ln |h|}{h} = \lim_{h\to 0} \ln |h|,$$

is infinite, and the derivative does not exist.

c. If $f(x) = x/\ln |x|$, then

$$f'(0) = \lim_{h\to 0} \frac{h}{h \ln |h|} = \lim_{h\to 0} \frac{1}{\ln |h|} = 0,$$

so the derivative does exist. But

$$f(0+h) - f(0) - hf'(0) = \frac{h}{\ln |h|}$$

is larger than $h^2$, since the limit

$$\lim_{h\to 0} \frac{h}{h^2 \ln |h|} = \lim_{h\to 0} \frac{1}{h \ln |h|}$$

is infinite: the denominator tends to 0 as $h$ tends to 0.

**1.7.5** a. Compute the partial derivatives:

$$D_1 f \begin{pmatrix} x \\ y \end{pmatrix} = \frac{x}{\sqrt{x^2+y}} \quad \text{and} \quad D_2 f \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2\sqrt{x^2+y}}.$$

This gives

$$D_1 f \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{2}{\sqrt{2^2+1}} = \frac{2}{\sqrt{5}} \quad \text{and} \quad D_2 f \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{1}{2\sqrt{2^2+1}} = \frac{1}{2\sqrt{5}}.$$

At the point $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$, we have $x^2 + y < 0$, so the function is not defined there, and neither are the partial derivatives.

b. Similarly, $D_1 f \begin{pmatrix} x \\ y \end{pmatrix} = 2xy$ and $D_2 f \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + 4y^3$. This gives

$$D_1 f \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 4 \quad \text{and} \quad D_2 f \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 4 + 4 = 8;$$

$$D_1 f \begin{pmatrix} 1 \\ -2 \end{pmatrix} = -4 \quad \text{and} \quad D_2 f \begin{pmatrix} 1 \\ -2 \end{pmatrix} = 1 + 4 \cdot (-8) = -31.$$

c. Compute

$$D_1 f\begin{pmatrix} x \\ y \end{pmatrix} = -y \sin xy$$
$$D_2 f\begin{pmatrix} x \\ y \end{pmatrix} = -x \sin xy + \cos y - y \sin y.$$

This gives

$$D_1 f\begin{pmatrix} 2 \\ 1 \end{pmatrix} = -\sin 2 \quad \text{and} \quad D_2 f\begin{pmatrix} 2 \\ 1 \end{pmatrix} = -2\sin 2 + \cos 1 - \sin 1$$
$$D_1 f\begin{pmatrix} 1 \\ -2 \end{pmatrix} = -2\sin 2 \quad \text{and} \quad D_2 f\begin{pmatrix} 1 \\ -2 \end{pmatrix} = \sin 2 + \cos 2 - 2\sin 2 = \cos 2 - \sin 2$$

d. Since

$$D_1 f\begin{pmatrix} x \\ y \end{pmatrix} = \frac{xy^2 + 2y^4}{2(x+y^2)^{3/2}} \quad \text{and} \quad D_2 f\begin{pmatrix} x \\ y \end{pmatrix} = \frac{2x^2y + xy^3}{(x+y^2)^{3/2}},$$

we have

$$D_1 f\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{4}{2\sqrt{27}} \quad \text{and} \quad D_2 f\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{10}{\sqrt{27}};$$
$$D_1 f\begin{pmatrix} 1 \\ -2 \end{pmatrix} = \frac{36}{10\sqrt{5}} \quad \text{and} \quad D_2 f\begin{pmatrix} 1 \\ -2 \end{pmatrix} = -\frac{12}{5\sqrt{5}}.$$

**1.7.6** a. We have

$$\frac{\partial \vec{\mathbf{f}}}{\partial x}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} -\sin x \\ 2xy \\ 2x\cos(x^2 - y) \end{bmatrix} \quad \text{and} \quad \frac{\partial \vec{\mathbf{f}}}{\partial y}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} 0 \\ x^2 + 2y \\ -\cos(x^2 - y) \end{bmatrix}.$$

b. Similarly,

$$\frac{\partial \vec{\mathbf{f}}}{\partial x}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} \frac{x}{\sqrt{x^2+y^2}} \\ y \\ 2y \sin xy \ \cos xy \end{bmatrix} \quad \text{and} \quad \frac{\partial \vec{\mathbf{f}}}{\partial y}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} \frac{y}{\sqrt{x^2+y^2}} \\ x \\ 2x \sin xy \ \cos xy \end{bmatrix}.$$

**1.7.7** Just pile up the partial derivative vectors side by side:

a. $$\left[\mathbf{D}\vec{\mathbf{f}}\begin{pmatrix} x \\ y \end{pmatrix}\right] = \begin{bmatrix} -\sin x & 0 \\ 2xy & x^2 + 2y \\ 2x\cos(x^2 - y) & -\cos(x^2 - y) \end{bmatrix}$$

b. $$\left[\mathbf{D}\vec{\mathbf{f}}\begin{pmatrix} x \\ y \end{pmatrix}\right] = \begin{bmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ y & x \\ 2y \sin xy \ \cos xy & 2x \sin xy \ \cos xy \end{bmatrix}.$$

**1.7.8** a. $D_1 f_1 = 2x\cos(x^2 + y)$, $D_2 f_1 = \cos(x^2 + y)$, $D_2 f_2 = xe^{xy}$

b. $3 \times 2$.

**1.7.9** a. The derivative is an $m \times n$ matrix

b. a $1 \times 3$ matrix (line matrix)

c. a $4 \times 1$ matrix (vector 4 high)

**1.7.10** a. Since $\mathbf{f}$ is linear, $\mathbf{f}(\mathbf{a}+\vec{\mathbf{v}}) = \mathbf{f}(\mathbf{a}) + \mathbf{f}(\mathbf{v})$. But since $\mathbf{f}$ is linear, $\mathbf{f}(\mathbf{v}) = [\mathbf{Df}(\mathbf{a})]\vec{\mathbf{v}}$:

$$\lim_{\vec{\mathbf{h}}\to 0} \frac{1}{|\vec{\mathbf{h}}|}\big(\mathbf{f}(\mathbf{a}+\vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a}) - \mathbf{f}(\vec{\mathbf{h}})\big) = \mathbf{0}, \quad \text{so} \quad [\mathbf{Df}(\mathbf{a})]\vec{\mathbf{h}} = \mathbf{f}(\vec{\mathbf{h}}).$$

b. The claim that $[\mathbf{Df}(\mathbf{a})]\vec{\mathbf{v}} = \mathbf{f}(\mathbf{a}+\vec{\mathbf{v}}) - \mathbf{f}(\mathbf{a})$ contradicts the definition of derivative:

$$[\mathbf{Df}(\mathbf{a})]\vec{\mathbf{v}} = \lim_{\vec{\mathbf{v}}\to 0} \frac{\mathbf{f}(\mathbf{a}+\vec{\mathbf{v}}) - \mathbf{f}(\mathbf{a})}{|\vec{\mathbf{v}}|}.$$

**1.7.11**

a. $\big[y\cos(xy),\, x\cos(xy)\big]$ b. $\big[2xe^{x^2+y^3},\, 3y^2e^{x^2+y^3}\big]$

c. $\begin{bmatrix} y & x \\ 1 & 1 \end{bmatrix}$ d. $\begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$

**1.7.12**

**1.7.13** For the first part, $|x|$ and $mx$ are continuous functions, hence so is $f(0+h) - f(0) - mh = |h| - mh$.

For the second, we have

$$\frac{|h| - mh}{h} = \frac{-h - mh}{h} = -1 - m \quad \text{when } h < 0$$
$$\frac{|h| - mh}{h} = \frac{h - mh}{h} = 1 - m \quad \text{when } h > 0.$$

The difference between these values is always 2, and cannot be made small by taking $h$ small.

**1.7.14** Since $\mathbf{g}$ is differentiable at $\mathbf{a}$,

$$\lim_{\vec{\mathbf{h}}\to\mathbf{0}} \frac{\mathbf{g}(\mathbf{a}+\vec{\mathbf{h}}) - \mathbf{g}(\mathbf{a}) - [\mathbf{Dg}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|} = \mathbf{0}.$$

This means that for every $\epsilon > 0$, there exists $\delta$ such that if $0 < |\vec{\mathbf{h}}| < \delta$, then

$$\left|\frac{\mathbf{g}(\mathbf{a}+\vec{\mathbf{h}}) - \mathbf{g}(\mathbf{a}) - [\mathbf{Dg}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right| \le \epsilon,$$

The triangle inequality (first inequality below) and proposition 1.4.11 (second inequality) then give

$$\left|\frac{\mathbf{g}(\mathbf{a}+\vec{\mathbf{h}}) - \mathbf{g}(\mathbf{a})}{|\vec{\mathbf{h}}|}\right| \le \left|[\mathbf{Dg}(\mathbf{a})]\frac{\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right| + \epsilon \le \big|[\mathbf{Dg}(\mathbf{a})]\big| \left|\frac{\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right| + \epsilon = \big|[\mathbf{Dg}(\mathbf{a})]\big| + \epsilon.$$

Solution 1.7.15, part a: The absolute value in the numerator is optional (but not in the denominator: you cannot divide by matrices).

Since $H$ is an $n \times m$ matrix, the $[0]$ in $\lim_{H\to[0]}$ is the $n \times m$ matrix with all entries 0.

**1.7.15** a. There exists a linear transformation $[\mathbf{D}F(A)]$ such that

$$\lim_{H\to[0]} \frac{\left|F(A+H) - F(A) - [\mathbf{D}F(A)]H\right|}{|H|} = 0.$$

b. The derivative is $[\mathbf{D}F(A)]H = AH^\top + HA^\top$. We found this by looking for linear terms in $H$ of the difference

$$\begin{aligned} F(A+H) - F(A) &= (A+H)(A+H)^\top - AA^\top \\ &= (A+H)(A^\top + H^\top) - AA^\top \\ &= AH^\top + HA^\top + HH^\top; \end{aligned}$$

see remark 1.7.6. The linear terms $AH^\top + HA^\top$ are the derivative. Indeed,

$$\begin{aligned} &\lim_{H\to[0]} \frac{|(A+H)(A+H)^\top - AA^\top - AH^\top - HA^\top|}{|H|} \\ &= \lim_{H\to[0]} \frac{|HH^\top|}{|H|} \le \lim_{H\to[0]} \frac{|H|\,|H^\top|}{|H|} = \lim_{H\to[0]} |H| = 0. \end{aligned}$$

**1.7.16** a. As a mapping $\mathbb{R}^4 \to \mathbb{R}^4$, the mapping $S$ is given by

$$S\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} a^2 + bc \\ ab + bd \\ ac + cd \\ bc + d^2 \end{pmatrix}.$$

b. The derivative of $S$ is given by the Jacobian matrix

$$\left[\mathbf{D}S\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}\right] = \begin{bmatrix} 2a & c & b & 0 \\ b & a+d & 0 & b \\ c & 0 & a+d & c \\ 0 & c & b & 2d \end{bmatrix}.$$

c. Let $B = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$. Then

$$\begin{aligned} &\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} + \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \begin{bmatrix} ax_1 + bx_3 & ax_2 + bx_4 \\ cx_1 + dx_3 & cx_2 + dx_4 \end{bmatrix} + \begin{bmatrix} ax_1 + cx_2 & bx_1 + dx_2 \\ ax_3 + cx_4 & bx_3 + dx_4 \end{bmatrix} \\ &= \begin{bmatrix} 2ax_1 + cx_2 + bx_3 & bx_1 + (a+d)x_2 + bx_4 \\ cx_1 + (a+d)x_3 + cx_4 & cx_2 + bx_3 + 2dx_4 \end{bmatrix}. \end{aligned}$$

It is indeed true that

$$\begin{bmatrix} 2ax_1 + cx_2 + bx_3 \\ bx_1 + (a+d)x_2 + bx_4 \\ cx_1 + (a+d)x_3 + cx_4 \\ cx_2 + bx_3 + 2dx_4 \end{bmatrix} = \begin{bmatrix} 2a & c & b & 0 \\ b & a+d & 0 & b \\ c & 0 & a+d & c \\ 0 & c & b & 2d \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

d. First compute the square of a $3 \times 3$ matrix $A$:

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}^2 = \begin{bmatrix} a_1^2 + b_1a_2 + c_1a_3 & a_1b_1 + b_1b_2 + c_1b_3 & a_1c_1 + b_1c_2 + c_1c_3 \\ a_2a_1 + b_2a_2 + c_2a_3 & a_2b_1 + b_2^2 + c_2b_3 & a_2c_1 + b_2c_2 + c_2c_3 \\ a_3a_1 + b_3a_2 + c_3a_3 & a_3b_1 + b_3b_2 + c_3b_3 & a_3c_1 + b_3c_2 + c_3^2 \end{bmatrix}.$$

This can be thought of as the mapping

$$S : \begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \\ a_3 \\ b_3 \\ c_3 \end{bmatrix} \mapsto \begin{bmatrix} a_1^2 + b_1a_2 + c_1a_3 \\ a_1b_1 + b_1b_2 + c_1b_3 \\ a_1c_1 + b_1c_2 + c_1c_3 \\ a_2a_1 + b_2a_2 + c_2a_3 \\ a_2b_1 + b_2^2 + c_2b_3 \\ a_2c_1 + b_2c_2 + c_2c_3 \\ a_3a_1 + b_3a_2 + c_3a_3 \\ a_3b_1 + b_3b_2 + c_3b_3 \\ a_3c_1 + b_3c_2 + c_3^2 \end{bmatrix}$$

with Jacobian matrix

$$[\mathbf{D}S(A)] = \begin{bmatrix} 2a_1 & a_2 & a_3 & b_1 & 0 & 0 & c_1 & 0 & 0 \\ b_1 & a_1 + b_2 & b_3 & 0 & b_1 & 0 & 0 & c_1 & 0 \\ c_1 & c_2 & a_1 + c_3 & 0 & 0 & b_1 & 0 & 0 & c_1 \\ a_2 & 0 & 0 & a_1 + b_2 & a_2 & a_3 & c_2 & 0 & 0 \\ 0 & a_2 & 0 & b_1 & 2b_2 & b_3 & 0 & c_2 & 0 \\ 0 & 0 & a_2 & c_1 & c_2 & b_2 + c_3 & 0 & 0 & c_2 \\ a_3 & 0 & 0 & b_3 & 0 & 0 & a_1 + c_3 & a_2 & a_3 \\ 0 & a_3 & 0 & 0 & b_3 & 0 & b_1 & b_2 + c_3 & b_3 \\ 0 & 0 & a_3 & 0 & 0 & b_3 & c_1 & c_2 & 2c_3 \end{bmatrix}.$$

Now compute $XA + AX$:

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 \end{bmatrix} + \begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 \end{bmatrix} \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} =$$

$$\begin{bmatrix} a_1x_1+a_2x_2+a_3x_3+a_1x_1+b_1x_4+c_1x_7 & b_1x_1+b_2x_2+b_3x_3+a_1x_2+b_1x_5+c_1x_8 & c_1x_1+c_2x_2+c_3x_3+a_1x_3+b_1x_6+c_1x_9 \\ a_1x_4+a_2x_5+a_3x_6+a_2x_1+b_2x_4+c_2x_7 & b_1x_4+b_2x_5+b_3x_6+a_2x_2+b_2x_5+c_2x_8 & c_1x_4+c_2x_5+c_3x_6+a_2x_3+b_2x_6+c_2x_9 \\ a_1x_7+a_2x_8+a_3x_9+a_3x_1+b_3x_4+c_3x_7 & b_1x_7+b_2x_8+b_3x_9+a_3x_2+b_3x_5+c_3x_8 & c_1x_7+c_2x_8+c_3x_9+a_3x_3+b_3x_6+c_3x_9 \end{bmatrix}.$$

Indeed, this is the same as $[\mathbf{D}S(A)]\mathbf{x}$.

Solution 1.7.17: This is sort of a miracle; the expressions should not be equal, they should differ by terms in $\epsilon^2$. The reason why they are exactly equal here is that

$$\begin{bmatrix} 0 & 0 \\ \epsilon & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

**1.7.17** The derivative of the squaring function is given by

$$[\mathbf{D}S(A)]H = AH + HA;$$

substituting $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $H = \begin{bmatrix} 0 & 0 \\ \epsilon & 0 \end{bmatrix}$ gives

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ \epsilon & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \epsilon & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \epsilon & 0 \\ \epsilon & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \epsilon & \epsilon \end{bmatrix} = \begin{bmatrix} \epsilon & 0 \\ 2\epsilon & \epsilon \end{bmatrix}.$$

Computing $(A+H)^2 - A^2$ gives the same result;

$$(A+H)^2 - A^2 = \begin{bmatrix} 1+\epsilon & 2 \\ 2\epsilon & 1+\epsilon \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \epsilon & 0 \\ 2\epsilon & \epsilon \end{bmatrix}.$$

**1.7.18** In the case of $2 \times 2$ matrices we have

$$S(A) : \begin{bmatrix} a & b \\ c & d \end{bmatrix} \to \begin{bmatrix} a^2+bc & b(a+d) \\ c(a+d) & bc+d^2 \end{bmatrix}.$$

Considering the elements of a $2 \times 2$ matrix to form a vector in $\mathbb{R}^4$ (ordered $a, b, c, d$) we see that the Jacobian of $S$ is:

$$\begin{bmatrix} 2a & c & b & 0 \\ b & a+d & 0 & b \\ c & 0 & a+d & c \\ 0 & c & b & 2d \end{bmatrix}$$

If $H$ is a matrix whose entries are 0 except for the $i^{th}$ one which is $h$ (using the above enumeration; e.g., if $i = 3$ we have $\begin{bmatrix} 0 & 0 \\ h & 0 \end{bmatrix}$), then $AH + HA$ is the matrix equal to $h$ times the $i$th column of the Jacobian.

**1.7.19** Since $\lim_{\vec{\mathbf{h}} \to \mathbf{0}} \frac{|\vec{\mathbf{h}}|\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|} = \mathbf{0}$, the derivative exists at the origin and is the 0 linear transformation, represented by the $n \times n$ matrix with all entries 0.

**1.7.20** The derivative is

$$\frac{1}{(ad-bc)^2} \begin{bmatrix} +d^2 & -cd & -db & +bc \\ -bd & +ad & +b^2 & -ab \\ -dc & +c^2 & +ad & -ac \\ bc & -ac & -ab & a^2 \end{bmatrix}.$$

This is obtained first by computing the inverse of $A$:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Then one computes

$$-\frac{1}{(ad-bc)^2} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

computing the matrix multiplication part as follows:

$$\begin{array}{ccc} & \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} & \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \\ \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} & \begin{bmatrix} dx_1 - bx_3 & dx_2 - bx_4 \\ -cx_1 + ax_3 & -cx_2 + ax_4 \end{bmatrix} & \begin{bmatrix} d^2x_1 - dbx_3 - cdx_2 + bcx_4 & -bdx_1 + b^2x_3 + adx_1 - abx_4 \\ -dcx_1 + adx_3 + c^2x_2 - acx_4 & bcx - -1 - abx_3 - acd_2 + a^2x_4 \end{bmatrix}. \end{array}$$

**1.7.21** We will work directly from the definition of the derivative:

$$\begin{aligned} \det(I+H) - \det(I) - (h_{1,1} + h_{2,2}) \\ &= (1+h_{1,1})(1+h_{2,2}) - h_{1,2}h_{2,1} - 1 - (h_{1,1}+h_{2,2}) \\ &= h_{1,1}h_{2,2} - h_{1,2}h_{2,1}. \end{aligned}$$

Each $h_{i,j}$ satisfies $|h_{i,j}| \le |H|$, so we have

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

Thus

$$\lim_{H\to 0} \frac{|\det(I+H) - \det(I) - (h_{1,1}+h_{2,2})|}{|H|} \le \lim_{H\to 0} 2|H| = 0.$$

**1.7.22**

**1.8.1** Three make sense:

c. $\mathbf{g}\circ\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$; the derivative is a $2\times 2$ matrix

d. $\mathbf{f}\circ\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^3$; the derivative is a $3\times 3$ matrix

e. $f\circ\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}$; the derivative is a $1\times 2$ matrix

**1.8.2** a. The derivative of $f$ at $\begin{pmatrix} a\\ b\\ c\end{pmatrix}$ is $[2a\ 2b\ 4c]$; the derivative $[\mathbf{Dg}]$ at $f\begin{pmatrix} a\\ b\\ c\end{pmatrix} = a^2+b^2+2c^2$ is $\begin{bmatrix} 1\\ 2(a^2+b^2+2c^2)\\ 3(a^2+b^2+2c^2)^2\end{bmatrix}$, so

$$\left[\mathbf{D}(\mathbf{g}\circ f)\begin{pmatrix} a\\ b\\ c\end{pmatrix}\right] = \begin{bmatrix} 1\\ 2(a^2+b^2+2c^2)\\ 3(a^2+b^2+2c^2)^2\end{bmatrix}[2a\,,2b\,,4c]$$
$$= \begin{bmatrix} 2a & 2b & 4c\\ 4a(a^2+b^2+2c^2) & 4b(a^2+b^2+2c^2) & 8c(a^2+b^2+2c^2)\\ 6a(a^2+b^2+2c^2)^2 & 6b(a^2+b^2+2c^2)^2 & 12c(a^2+b^2+2c^2)^2\end{bmatrix}.$$

b. The derivative of $\mathbf{f}$ at $\begin{pmatrix} x\\ y\\ z\end{pmatrix}$ is $\begin{bmatrix} 2x & 0 & 1\\ 0 & z & y\end{bmatrix}$, the derivative of $g$ at $\begin{bmatrix} a\\ b\end{bmatrix}$ is $[2a \quad 2b]$, and the derivative of $g$ at $\mathbf{f}\begin{pmatrix} x\\ y\\ z\end{pmatrix}$ is $[2x^2+2z \quad 2yz]$, so the derivative of $g\circ\mathbf{f}$ at $\begin{pmatrix} x\\ y\\ z\end{pmatrix}$ is

$$[2x^2+2z \quad 2yz]\begin{bmatrix} 2x & 0 & 1\\ 0 & z & y\end{bmatrix} = [4x^3+4xz \quad 2yz^2 \quad 2x^2+2z+2y^2z].$$

**1.8.3** Yes: We have a composition of sine, the exponential function, and the function $\begin{pmatrix} x\\ y\end{pmatrix} \mapsto xy$, all of which are differentiable everywhere.

**1.8.4** a. The following compositions exist:

(i) $\mathbf{f}\circ g : \mathbb{R}^2 \to \mathbb{R}^3$; (ii) $f\circ\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}$; (iii) $\mathbf{f}\circ f : \mathbb{R}^3 \to \mathbb{R}^3$; (iv) $f\circ\mathbf{f} : \mathbb{R} \to \mathbb{R}$.

(One could make more by using three functions; for example, $f\circ\mathbf{f}\circ g : \mathbb{R}^2 \to \mathbb{R}$.)

b. We have

$$\text{(i) } (\mathbf{f}\circ g)\begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{f}(2a+b^2) = \begin{pmatrix} 2a+b^2 \\ 4a+2b^2 \\ (2a+b^2)^2 \end{pmatrix} \quad \text{(ii) } (f\circ \mathbf{g})\begin{pmatrix} x \\ y \end{pmatrix} = f\begin{pmatrix} \cos x \\ x+y \\ \sin y \end{pmatrix} = \cos^2 x + (x+y)^2.$$

$$\text{(iii) } (\mathbf{f}\circ f)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{f}(x^2+y^2) = \begin{pmatrix} x^2+y^2 \\ 2(x^2+y^2) \\ (x^2+y^2)^2 \end{pmatrix} \quad \text{(iv) } (f\circ \mathbf{f})(t) = f\begin{pmatrix} t \\ 2t \\ t^2 \end{pmatrix} = t^2+4t^2 = 5t^2.$$

c.

(i) Computing the derivative directly from $(\mathbf{f}\circ g)\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2a+b^2 \\ 4a+2b^2 \\ (2a+b^2)^2 \end{pmatrix}$ gives $\left[\mathbf{D}(\mathbf{f}\circ g)\begin{pmatrix} a \\ b \end{pmatrix}\right] = \begin{bmatrix} 2 & 2b \\ 4 & 4b \\ 8a+4b^2 & 8ab+4b^3 \end{bmatrix}$; since $[\mathbf{Df}(t)] = \begin{bmatrix} 1 \\ 2 \\ 2t \end{bmatrix}$ and $\left[\mathbf{D}g\begin{pmatrix} a \\ b \end{pmatrix}\right] = [2 \quad 2b]$, the chain rule gives

$$\left[\mathbf{Df}\left(g\begin{pmatrix} a \\ b \end{pmatrix}\right)\right]\left[\mathbf{D}g\begin{pmatrix} a \\ b \end{pmatrix}\right] = \begin{bmatrix} 1 \\ 2 \\ 4a+2b^2 \end{bmatrix}[2 \quad 2b] = \begin{bmatrix} 2 & 2b \\ 4 & 4b \\ 8a+4b^2 & 8ab+4b^3 \end{bmatrix}.$$

(ii) Computing the derivative directly from

$$(f\circ \mathbf{g})\begin{pmatrix} x \\ y \end{pmatrix} = \cos^2 x + (x+y)^2 = \cos^2 x + x^2 + 2xy + y^2$$

gives $\left[\mathbf{D}(f\circ \mathbf{g})\begin{pmatrix} x \\ y \end{pmatrix}\right] = \begin{bmatrix} -2\cos x \sin x + 2x + 2y & 2x+2y \end{bmatrix}$; since

$$\left[\mathbf{D}f\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right] = [2x \quad 2y \quad 0],$$

$$\left[\mathbf{D}f\left(\mathbf{g}\begin{pmatrix} x \\ y \end{pmatrix}\right)\right] = \left[\mathbf{D}f\begin{pmatrix} \cos x \\ x+y \\ \sin y \end{pmatrix}\right] = [2\cos x \quad 2x+2y \quad 0],$$

$$\left[\mathbf{Dg}\begin{pmatrix} x \\ y \end{pmatrix}\right] = \begin{bmatrix} -\sin x & 0 \\ 1 & 1 \\ 0 & \cos y \end{bmatrix},$$

the chain rule gives

$$[2\cos x \quad 2x+2y \quad 0]\begin{bmatrix} -\sin x & 0 \\ 1 & 1 \\ 0 & \cos y \end{bmatrix} = \begin{bmatrix} -2\cos x \sin x + 2x + 2y & 2x+2y \end{bmatrix}.$$

(iii) Computing the derivative from $(\mathbf{f}\circ f)\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}x^2+y^2\\2(x^2+y^2)\\(x^2+y^2)^2\end{pmatrix}$ gives $\left[\mathbf{D}(\mathbf{f}\circ f)\begin{pmatrix}x\\y\\z\end{pmatrix}\right]=\begin{bmatrix}2x & 2y & 0\\4x & 4y & 0\\4x^3+4xy^2 & 4x^2y+4y^3 & 0\end{bmatrix}$; the chain rule gives

$$\left[\mathbf{Df}\left(f\begin{pmatrix}x\\y\\z\end{pmatrix}\right)\right]\left[\mathbf{D}f\begin{pmatrix}x\\y\\z\end{pmatrix}\right]=\begin{bmatrix}1\\2\\2x^2+2y^2\end{bmatrix}[2x\quad 2y\quad 0]$$
$$=\begin{bmatrix}2x & 2y & 0\\4x & 4y & 0\\4x^3+4xy^2 & 4x^2y+4y^3 & 0\end{bmatrix}.$$

(iv) Computing the derivative directly from $(f\circ\mathbf{f})(t)=t^2+4t^2=5t^2$ gives $[\mathbf{D}((f\circ\mathbf{f}))(t)]=10t$; the chain rule gives

$$[2t\quad 4t\quad 0]\begin{bmatrix}1\\2\\2t\end{bmatrix}=2t+8t=10t.$$

**1.8.5** One must also show that $fg$ is differentiable, working from the definition of the derivative.

**1.8.6** a. We need to prove that

$$\lim_{\vec{\mathbf{h}}\to\mathbf{0}}\frac{\left|\mathbf{f}(\mathbf{a}+\vec{\mathbf{h}})\cdot\mathbf{g}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{f}(\mathbf{a})\cdot\mathbf{g}(\mathbf{a})-\mathbf{f}(\mathbf{a})\cdot\Big([\mathbf{Dg}(\mathbf{a})]\vec{\mathbf{h}}\Big)-\Big([\mathbf{Df}(\mathbf{a})]\vec{\mathbf{h}}\Big)\cdot\mathbf{g}(\mathbf{a})\right|}{|\vec{\mathbf{h}}|}=\mathbf{0}.$$

Since the term under the limit can be written

$$\Big(\mathbf{f}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{f}(\mathbf{a})\Big)\cdot\frac{\mathbf{g}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{g}(\mathbf{a})}{|\vec{\mathbf{h}}|}+\mathbf{f}(\mathbf{a})\cdot\left(\frac{\mathbf{g}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{g}(\mathbf{a})-[\mathbf{Dg}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right)$$
$$+\left(\frac{\mathbf{f}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{f}(\mathbf{a})-[\mathbf{Df}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right)\cdot\mathbf{g}(\mathbf{a}),$$

it is enough to prove that the three limits

$$\lim_{\vec{\mathbf{h}}\to 0}\left|\mathbf{f}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{f}(\mathbf{a})\right|\frac{|\mathbf{g}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{g}(\mathbf{a})|}{|\vec{\mathbf{h}}|}$$
$$\lim_{\vec{\mathbf{h}}\to 0}|\mathbf{f}(\mathbf{a})|\left|\frac{\mathbf{g}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{g}(\mathbf{a})-[\mathbf{Dg}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right|$$
$$\lim_{\vec{\mathbf{h}}\to 0}\left|\frac{\mathbf{f}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{f}(\mathbf{a})-[\mathbf{Df}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right||\mathbf{g}(\mathbf{a})|$$

all vanish.

The first vanishes because

$$\frac{|\mathbf{g}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{g}(\mathbf{a})|}{|\vec{\mathbf{h}}|}$$

is bounded when $\vec{\mathbf{h}} \to \mathbf{0}$, and the factor $\left|f(\mathbf{a}+\vec{\mathbf{h}})-f(\mathbf{a})\right|$ tends to $\mathbf{0}$. The second vanishes because

$$\left|\frac{\mathbf{g}(\mathbf{a}+\vec{\mathbf{h}})-\mathbf{g}(\mathbf{a})-[\mathbf{Dg}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right|$$

tends to 0 when $\vec{\mathbf{h}} \to \mathbf{0}$, and the factor $f(\mathbf{a})|$ is constant. The third vanishes because

$$\left|\frac{f(\mathbf{a}+\vec{\mathbf{h}})-f(\mathbf{a})-[\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right|$$

tends to 0 as $\vec{\mathbf{h}} \to \mathbf{0}$, and the factor $|\mathbf{g}(\mathbf{a})|$ is constant.

b. The derivative is given by the formula

$$[\mathbf{D}(\vec{\mathbf{f}}\times\vec{\mathbf{g}})(\mathbf{a})]\vec{\mathbf{h}} = \Big(\big([\mathbf{D}\vec{\mathbf{f}}(\mathbf{a})]\vec{\mathbf{h}}\big)\times\vec{\mathbf{g}}(\mathbf{a})\Big) + \Big(\vec{\mathbf{f}}(\mathbf{a})\times\big([\mathbf{D}\vec{\mathbf{g}}(\mathbf{a})]\vec{\mathbf{h}}\big)\Big).$$

The proof that this is correct is again almost identical to part a or b. We need to prove that

$$\lim_{\vec{\mathbf{h}}\to\mathbf{0}} \frac{\Big|\Big(\vec{\mathbf{f}}(\mathbf{a}+\vec{\mathbf{h}})\times\vec{\mathbf{g}}(\mathbf{a}+(\vec{\mathbf{h}})\Big)-\Big(\vec{\mathbf{f}}(\mathbf{a})\times\vec{\mathbf{g}}(\mathbf{a})\Big)-\Big(\vec{\mathbf{f}}(\mathbf{a})\times\big([\mathbf{D}\vec{\mathbf{g}}(\mathbf{a})]\vec{\mathbf{h}}\big)\Big)-\Big(\big([\mathbf{D}\vec{\mathbf{f}}(\mathbf{a})]\vec{\mathbf{h}}\big)\times\vec{\mathbf{g}}(\mathbf{a})\Big)\Big|}{|\vec{\mathbf{h}}|} = \mathbf{0}.$$

The term under the limit can be written as the sum of three cross products:

$$\Big(\vec{\mathbf{f}}(\mathbf{a}+\vec{\mathbf{h}})-\vec{\mathbf{f}}(\mathbf{a})\Big)\times\frac{\vec{\mathbf{g}}(\mathbf{a}+\vec{\mathbf{h}})-\vec{\mathbf{g}}(\mathbf{a})}{|\vec{\mathbf{h}}|} \quad + \quad \vec{\mathbf{f}}(\mathbf{a})\times\left(\frac{\vec{\mathbf{g}}(\mathbf{a}+\vec{\mathbf{h}})-\vec{\mathbf{g}}(\mathbf{a})-[\mathbf{D}\vec{\mathbf{g}}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right)$$
$$+\left(\frac{\vec{\mathbf{f}}(\mathbf{a}+\vec{\mathbf{h}})-\vec{\mathbf{f}}(\mathbf{a})-[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right)\times\vec{\mathbf{g}}(\mathbf{a}),$$

and, since the area of a parallelogram is at most the product of the lengths of the sides, we have

$$|\vec{\mathbf{f}}(\mathbf{x})\times\vec{\mathbf{g}}(\mathbf{x})| \le |\vec{\mathbf{f}}(\mathbf{x})|\,|\vec{\mathbf{g}}(\mathbf{x})|.$$

Thus it is enough to prove that the three limits

$$\lim_{\vec{\mathbf{h}}\to\mathbf{0}}\left|\vec{\mathbf{f}}(\mathbf{a}+\vec{\mathbf{h}})-\vec{\mathbf{f}}(\mathbf{a})\right|\frac{|\vec{\mathbf{g}}(\mathbf{a}+\vec{\mathbf{h}})-\vec{\mathbf{g}}(\mathbf{a})|}{|\vec{\mathbf{h}}|}$$
$$\lim_{\vec{\mathbf{h}}\to\mathbf{0}}|\vec{\mathbf{f}}(\mathbf{a})|\left|\frac{\vec{\mathbf{g}}(\mathbf{a}+\vec{\mathbf{h}})-\vec{\mathbf{g}}(\mathbf{a})-[\mathbf{D}\vec{\mathbf{g}}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right|$$
$$\lim_{\vec{\mathbf{h}}\to\mathbf{0}}\left|\frac{\vec{\mathbf{f}}(k\mathbf{a}+\vec{\mathbf{h}})-\vec{\mathbf{f}}(\mathbf{a})-[\mathbf{D}\vec{\mathbf{f}}(\mathbf{a})]\vec{\mathbf{h}}}{|\vec{\mathbf{h}}|}\right||\vec{\mathbf{g}}(\mathbf{a})|$$

all vanish, which again happens for the same reasons as in part a.

**1.8.7** Since $[\mathbf{D}f(x)] = [x_2 \quad x_1 + x_3 \quad x_2 + x_4 \quad \cdots \quad x_{n-2} + x_n \quad x_{n-1}]$, we have

$[\mathbf{D}\big(f(\gamma(t))\big)] = [t^2 \quad t + t^3 \quad t^2 + t^4 \quad \cdots \quad t^{n-2} + t^n \qquad t^{n-1}]$. In addition, $[\mathbf{D}\gamma(t)] = \begin{bmatrix} 1 \\ 2t \\ \vdots \\ nt^{n-1} \end{bmatrix}$. So the derivative of the function $t \to f(\gamma(t))$ is

$$\underbrace{[\mathbf{D}(f \circ \gamma)(t)]}_{\text{deriv. of comp. at } t} = \underbrace{[\mathbf{D}f(\gamma(t))]}_{\substack{\text{deriv. of } f \\ \text{at } \gamma(t)}} \underbrace{[\mathbf{D}\gamma(t)]}_{\text{deriv. of } \gamma \text{at } t} = t^2 + \left(\sum_{i=2}^{n-1} it^{i-1}(t^{i-1} + t^{i+1})\right) + nt^{2(n-1)}$$

**1.8.8** True. If there were such a mapping $\mathbf{g}$, then

$$\left[\mathbf{Df}\begin{pmatrix}0\\0\end{pmatrix}\right]\left[\mathbf{Dg}\begin{pmatrix}1\\1\end{pmatrix}\right] = \left[\mathbf{Df} \circ \mathbf{g}\begin{pmatrix}1\\1\end{pmatrix}\right] = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}.$$

The first equality is the chain rule, the second comes from the fact that $\begin{pmatrix}x\\y\end{pmatrix} \mapsto \begin{pmatrix}y\\x\end{pmatrix}$ is linear, so its derivative is itself.

So let $\left[\mathbf{Dg}\begin{pmatrix}1\\1\end{pmatrix}\right] = \begin{bmatrix}a & b\\c & d\end{bmatrix}$; our equation above says

$$\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}a & b\\c & d\end{bmatrix} = \begin{bmatrix}a+c & b+d\\a+c & b+d\end{bmatrix} = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}.$$

This equation has no solutions, since $a+c$ must simultaneously be 1 and 0.

Solution 1.8.9: This isn't really a good problem to test knowledge of the chain rule, because it is easiest to solve it without ever invoking the chain rule (at least in several variables).

**1.8.9** Clearly

$$D_1 f\begin{pmatrix}x\\y\end{pmatrix} = 2xy\varphi'(x^2-y^2); \qquad D_2 f\begin{pmatrix}x\\y\end{pmatrix} = -2y^2\varphi'(x^2-y^2) + \varphi(x^2-y^2).$$

Thus

$$\begin{aligned}\frac{1}{x}D_1 f\begin{pmatrix}x\\y\end{pmatrix} + \frac{1}{y}D_2 f\begin{pmatrix}x\\y\end{pmatrix} &= 2y\varphi'(x^2-y^2) - 2y\varphi'(x^2-y^2) + \frac{1}{y}\varphi(x^2-y^2)\\ &= \frac{1}{y^2}f\begin{pmatrix}x\\y\end{pmatrix}.\end{aligned}$$

To use the chain rule, write $f = k \circ \mathbf{h} \circ \mathbf{g}$, where

$$\mathbf{g}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}x^2-y^2\\y\end{pmatrix}, \qquad \mathbf{h}\begin{pmatrix}u\\v\end{pmatrix} = \begin{pmatrix}\varphi(u)\\v\end{pmatrix}, \qquad k\begin{pmatrix}s\\t\end{pmatrix} = st.$$

This leads to

$$\left[\mathbf{D}f\begin{pmatrix}x\\y\end{pmatrix}\right] = [t,s]\begin{bmatrix}\varphi'(u) & 0\\0 & 1\end{bmatrix}\begin{bmatrix}2x & -2y\\0 & 1\end{bmatrix} = [2xt\varphi'(u),\ -2yt\varphi'(u)+s].$$

Insert the values of the variables; you find

$$D_1 f\begin{pmatrix}x\\y\end{pmatrix} = 2xy\,\varphi'(x^2-y^2) \text{ and } D_2 f\begin{pmatrix}x\\y\end{pmatrix} = -2y^2\varphi'(x^2-y^2) + \varphi(x^2-y^2).$$

Now continue as above.

**1.8.10** In the first printing, the problem was misstated. The problem should say: "If $f : \mathbb{R}^2 \to \mathbb{R}$ can be written … " (it should not be "If $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ can be written … "). The function $f$ must be scalar-valued, since $\varphi$ is.

(a) Let $f\begin{pmatrix} x \\ y \end{pmatrix} = \varphi(x^2+y^2)$. By the chain rule, we have

$$\left[\mathbf{D}f\begin{pmatrix} x \\ y \end{pmatrix}\right] = \left[D_1 f\begin{pmatrix} x \\ y \end{pmatrix}, D_2 f\begin{pmatrix} x \\ y \end{pmatrix}\right] = [\mathbf{D}\varphi(x^2+y^2)][2x, 2y].$$

So

$$D_1 f\begin{pmatrix} x \\ y \end{pmatrix} = 2x[\mathbf{D}\varphi(x^2+y^2)] \quad \text{and} \quad D_2 f\begin{pmatrix} x \\ y \end{pmatrix} = 2y[\mathbf{D}\varphi(x^2+y^2)].$$

The result follows immediately.

(b) Let $f$ satisfy $xD_2 f - yD_1 f = 0$, and let us show that it is constant on circles centered at the origin. This is the same thing as showing that for the function

$$g\begin{pmatrix} r \\ \theta \end{pmatrix} \stackrel{\text{def}}{=} f\begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix},$$

we have

$$D_\theta g = 0.$$

This derivative can be computed by the chain rule, to find

$$\begin{aligned} D_\theta g\begin{pmatrix} r \\ \theta \end{pmatrix} &= \left(D_1 f\begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}\right)(-r\sin\theta) + \left(D_2 f\begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}\right)(r\cos\theta) \\ &= xD_2 f\begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} - yD_1 f\begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} = 0. \end{aligned}$$

So $f\begin{pmatrix} x \\ y \end{pmatrix} = f\begin{pmatrix} \sqrt{x^2+y^2} \\ 0 \end{pmatrix}$, and we can take $\varphi(r) = f\begin{pmatrix} r \\ 0 \end{pmatrix}$.

**1.8.11** Using the chain rule in one variable,

$$D_1 f = \varphi'\left(\frac{x+y}{x-y}\right)\left(\frac{1(x-y)-1(x+y)}{(x-y)^2}\right) = \varphi'\left(\frac{x+y}{x-y}\right)\left(\frac{-2y}{(x-y)^2}\right)$$

and

$$D_2 f = \varphi'\left(\frac{x+y}{x-y}\right)\left(\frac{1(x-y)-(-1)(x+y)}{(x-y)^2}\right) = \varphi'\left(\frac{x+y}{x-y}\right)\left(\frac{2x}{(x-y)^2}\right)$$

so

$$xD_1 f + yD_2 f = \varphi'\left(\frac{x+y}{x-y}\right)\left(\frac{-2xy}{(x-y)^2}\right) + \varphi'\left(\frac{x+y}{x-y}\right)\left(\frac{2yx}{(x-y)^2}\right) = 0$$

**1.8.12** a. True: the chain rule tells us that

$$[\mathbf{D}(\mathbf{g}\circ\mathbf{f})(\mathbf{0})]\vec{\mathbf{h}} = [\mathbf{D}(\mathbf{g}(\mathbf{f}(\mathbf{0})))][\mathbf{D}\mathbf{f}(\mathbf{0})]\vec{\mathbf{h}}.$$

If there exists a differentiable function $\mathbf{g}$ such that $(\mathbf{g}\circ\mathbf{f})(\mathbf{x}) = \mathbf{x}$, then $[\mathbf{D}(\mathbf{g}\circ\mathbf{f})(\mathbf{0})] = I$ which would mean that

$$[\mathbf{D}(\mathbf{g}\circ\mathbf{f})(\mathbf{0})]\vec{\mathbf{h}} = [\mathbf{D}(\mathbf{g}(\mathbf{f}(\mathbf{0})))][\mathbf{D}\mathbf{f}(\mathbf{0})]\vec{\mathbf{h}} = \vec{\mathbf{h}},$$