MATH20132 Calculus of Several Variables. 2019-20

Solutions to Problems 5: C¹-functions and more

C^1-***scalar***-*valued functions*

1. Define the function $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(\mathbf{x}) = x \sin(xyz) + \exp(yz)$ where $\mathbf{x} = (x, y, z)^T$. Prove that f is a Fréchet differentiable function by showing that f is C^1 on $\mathbb{R}^3$.

Solution The partial derivatives of f are

$$\begin{aligned}
\frac{\partial f}{\partial x}(\mathbf{x}) &= \sin(xyz) + xyz\cos(xyz),\\
\frac{\partial f}{\partial y}(\mathbf{x}) &= x^2 z\cos(xyz) + z\exp(yz),\\
\frac{\partial f}{\partial z}(\mathbf{x}) &= x^2 y\cos(xyz) + y\exp(yz).
\end{aligned}$$

The xyz, x^2z, etc. terms are polynomials in the variables of $\mathbf{x}$ and so are continuous. The sin, cos, exp are functions from $\mathbb{R}$ to $\mathbb{R}$, known to be continuous from previous analysis courses. Hence, by the Composite Rule along with the Product and Sum Rules for continuous functions, the partial derivatives above are continuous on $\mathbb{R}^3$. Hence f is C^1 on $\mathbb{R}^3$ and thus Fréchet differentiable on $\mathbb{R}^3$.

2. Define the function $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(\mathbf{x}) = \sin(xy^2z^3)$ where $\mathbf{x} = (x, y, z)^T$.

i. Prove that f is Fréchet differentiable at $\mathbf{a} = (\pi, 1, -1)^T$.

ii. Find the directional derivative $d_{\mathbf{v}} f(\mathbf{a})$ where $\mathbf{v} = (2/3, 1/3, -2/3)^T$.

Solution i. For a general $\mathbf{x} \in \mathbb{R}^3$ the gradient vector is

$$\nabla f(\mathbf{x}) = \begin{pmatrix} y^2z^3\cos(xy^2z^3) \\ 2xyz^3\cos(xy^2z^3) \\ 3xy^2z^2\cos(xy^2z^3) \end{pmatrix}.$$

The components are continuous on $\mathbb{R}^3$ hence f is a C^1-function and thus Fréchet differentiable on $\mathbb{R}^3$ and hence at the given $\mathbf{a}$.

ii. It is **important** to make the observation that f is Fréchet differentiable because **only** if you know f is Fréchet differentiable at $\mathbf{a}$ can you say

$$\begin{aligned} d_{\mathbf{v}}f(\mathbf{a}) = \nabla f(\mathbf{a}) \bullet \mathbf{v} &= \frac{1}{3}\begin{pmatrix} -\cos(-\pi) \\ -2\pi\cos(-\pi) \\ 3\pi\cos(-\pi) \end{pmatrix} \bullet \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} \\ &= \frac{1}{3}\begin{pmatrix} 1 \\ 2\pi \\ -3\pi \end{pmatrix} \bullet \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} \\ &= \frac{2+8\pi}{3}. \end{aligned}$$

The following was Questions 1& 2 on Sheet 4 but now, with C^1-functions, we can give a quicker solution.

3. a. By using partial differentiation find the gradient vectors of

i. $f : \mathbb{R}^2 \to \mathbb{R}, \mathbf{x} \longmapsto x(x+y)$ and

ii. $g : \mathbb{R}^2 \to \mathbb{R}, \mathbf{x} \longmapsto y(x-y)$

and show they are everywhere Fréchet differentiable. Find the directional derivatives of f and g at $\mathbf{a} = (1,2)^T$ in the direction $\mathbf{v} = (2,-1)^T/\sqrt{5}$, justifying your method.

b. Using partial differentiation find the gradient vector of $h : \mathbb{R}^3 \to \mathbb{R}$ by $\mathbf{x} \longmapsto xy + yz + xz$ where $\mathbf{x} = (x,y,z)^T$, and show it is everywhere Fréchet differentiable. Find the directional derivative of f at $\mathbf{a} = (1,2,3)^T$ in the direction $\mathbf{v} = (3,2,1)^T/\sqrt{14}$, justifying your method.

Solution a. i. $\nabla f(\mathbf{x}) = (2x+y, x)^T$, ii. $\nabla g(\mathbf{x}) = (y, x-2y)^T$

All the terms of both gradient vectors are polynomials which are everywhere continuous hence both f and g are C^1-functions and thus everywhere Fréchet differentiable.

Since f and g are Fréchet differentiable we have

$$\begin{aligned} d_{\mathbf{v}}f(\mathbf{a}) &= \nabla f(\mathbf{a}) \bullet \mathbf{v} = \frac{1}{\sqrt{5}}\begin{pmatrix} 4 \\ 1 \end{pmatrix} \bullet \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \frac{7}{\sqrt{5}}, \\ d_{\mathbf{v}}g(\mathbf{a}) &= \nabla g(\mathbf{a}) \bullet \mathbf{v} = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -3 \end{pmatrix} \bullet \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \frac{7}{\sqrt{5}}. \end{aligned}$$

Hopefully they agree with your answers to Question 1 on Sheet 3.

b. $\nabla h(\mathbf{x}) = (y+z, x+z, x+y)^T$. All the terms of the gradient vector are polynomials which are everywhere continuous hence h is a C^1-function and thus everywhere Fréchet differentiable. Therefore we are allowed to say

$$d_{\mathbf{v}} h(\mathbf{a}) = \nabla h(\mathbf{a}) \cdot \mathbf{v} = \frac{1}{\sqrt{14}} \begin{pmatrix} 5 \\ 4 \\ 3 \end{pmatrix} \bullet \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \frac{26}{\sqrt{14}}.$$

Hopefully this agrees with your answer to Question 3 on Sheet 3.

4. (Tricky) *Recall:*

$$\boxed{f \text{ is } C^1 \text{ at } \mathbf{a} \implies f \text{ is Fréchet differentiable at } \mathbf{a} \implies f \text{ continuous at } \mathbf{a}.}$$

The contrapositive of this is

$$\boxed{f \text{ not conts at } \mathbf{a} \implies f \text{ not F-differentiable at } \mathbf{a} \implies f \text{ is not } C^1.} \quad (1)$$

Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(\mathbf{x}) = \frac{xy}{x^2+y^2} \quad \text{if } \mathbf{x} \neq \mathbf{0}; \quad \text{with } f(\mathbf{0}) = 0.$$

This was shown in Question 11ii on Sheet 1 to **not** be continuous at $\mathbf{0}$. So, as not to contradict (1), prove that f is **not** C^1 at $\mathbf{0}$, i.e. that the partial derivatives are not continuous at $\mathbf{0}$.

Solution Partial differentiation gives

$$\frac{\partial f}{\partial x}(\mathbf{x}) = \frac{y\left(y^2 - x^3\right)}{\left(x^2+y^2\right)^2}, \quad (2)$$

for $\mathbf{x} \neq \mathbf{0}$. Going back to the definition gives

$$\frac{\partial f}{\partial x}(\mathbf{0}) = \lim_{t \to 0} \frac{f(\mathbf{0} + t\mathbf{e}_1) - f(\mathbf{0})}{t} = \lim_{t \to 0} \frac{1}{t} \frac{t \times 0}{t^2 + 0^2} = 0.$$

To be C^1 at $\mathbf{0}$ means that $\partial f(\mathbf{x})/\partial x$ is continuous at $\mathbf{0}$. Look at the limit of $\partial f(\mathbf{x})/\partial x$ as $\mathbf{x} \to \mathbf{0}$ *along the* y*-axis*, i.e. $\mathbf{x} = t\mathbf{e}_2$ as $t \to 0$. For then, by (2),

$$\frac{\partial f}{\partial x}(t\mathbf{e}_2) = \frac{t^3}{t^4} = \frac{1}{t}$$

which has no limit as $t \to 0$ and certainly doesn't equal $0 = \partial f(\mathbf{0})/\partial x$. Hence the partial derivative w.r.t. x is not continuous.

The argument for the partial derivative w.r.t. y is identical but this is not needed; as soon as one partial derivative is not continuous we can deduce that f is not C^1.

*C^1-**vector**-valued functions*

5. Find the Jacobian matrices of the following functions, show that the functions are everywhere Fréchet differentiable and then find the directional derivatives at the given point $\mathbf{a}$ in the direction $\mathbf{v}$. In this way check your answers to Questions 5 & 7 on Sheet 3.

i. $\mathbf{f}\left((x,y,z)^T\right) = (xy, yz)^T$, $\mathbf{a} = (1,3,-2)^T$ and $\mathbf{v} = (-1,1,-2)^T/\sqrt{6}$,

ii. $\mathbf{f}\left((x,y)^T\right) = (xy^2, x^2y)^T$, $\mathbf{a} = (2,1)$ and $\mathbf{v} = (1,-1)^T/\sqrt{2}$.

Solution i. The Jacobian matrix is

$$J\mathbf{f}(\mathbf{a}) = \begin{pmatrix} y & x & 0 \\ 0 & z & y \end{pmatrix}_{\mathbf{x}=\mathbf{a}} = \begin{pmatrix} 3 & 1 & 0 \\ 0 & -2 & 3 \end{pmatrix}.$$

All the terms in $J\mathbf{f}(\mathbf{x})$ are polynomials and thus continuous on $\mathbb{R}^3$ and thus $\mathbf{f}$ is a C^1-function and hence everywhere Fréchet differentiable. Since, for a Fréchet differentiable function, $d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) = J\mathbf{f}(\mathbf{a})\,\mathbf{v}$ we have

$$d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) = \frac{1}{\sqrt{6}}\begin{pmatrix} 3 & 1 & 0 \\ 0 & -2 & 3 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix} = \frac{1}{\sqrt{6}}\begin{pmatrix} -2 \\ -8 \end{pmatrix} = -\sqrt{\frac{2}{3}}\begin{pmatrix} 1 \\ 4 \end{pmatrix}.$$

This agrees with Question 5 on Sheet 3.

ii. The Jacobian matrix is

$$J\mathbf{f}(\mathbf{a}) = \begin{pmatrix} y^2 & 2xy \\ 2xy & x^2 \end{pmatrix}_{\mathbf{x}=\mathbf{a}} = \begin{pmatrix} 1 & 4 \\ 4 & 4 \end{pmatrix}.$$

All the terms in $J\mathbf{f}(\mathbf{x})$ are polynomials and thus continuous on $\mathbb{R}^2$ and thus $\mathbf{f}$ is a C^1-function and hence everywhere Fréchet differentiable. Since, for a Fréchet differentiable function, $d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) = J\mathbf{f}(\mathbf{a})\,\mathbf{v}$ we have

$$d_{\mathbf{v}}\mathbf{f}(\mathbf{a}) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 4 \\ 4 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} -3 \\ 0 \end{pmatrix} = -\frac{3}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

This agrees with Question 7 on Sheet 3.

Chain Rule

6. Let

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} x^2 y \\ xy^2 \end{pmatrix} \quad \text{and} \quad \mathbf{g}(\mathbf{u}) = \begin{pmatrix} u+v \\ u-v \end{pmatrix},$$

for $\mathbf{x} = (x, y)^T$ and $\mathbf{u} = (u, v)^T$.

i. Calculate $\mathbf{f}\big(\mathbf{g}(\mathbf{u})\big)$ and thus find the Jacobian matrix $J(\mathbf{f} \circ \mathbf{g})(\mathbf{a})$ where $\mathbf{a} = (1, -2)^T$.

ii. Alternatively find $J\mathbf{f}(\mathbf{b})$, with $\mathbf{b} = \mathbf{g}(\mathbf{a})$, and $J\mathbf{g}(\mathbf{a})$ and use the Chain Rule to calculate $J(\mathbf{f} \circ \mathbf{g})(\mathbf{a})$

Solution i The composition function $\mathbf{f} \circ \mathbf{g}$ is

$$\mathbf{f}\big(\mathbf{g}(\mathbf{u})\big) = \begin{pmatrix} (u+v)^2 (u-v) \\ (u+v)(u-v)^2 \end{pmatrix} = \begin{pmatrix} u^3 + u^2 v - uv^2 - v^3 \\ u^3 - u^2 v - uv^2 + v^3 \end{pmatrix}$$

Thus

$$\begin{aligned} J(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) &= \begin{pmatrix} 3u^2 + 2uv - v^2 & u^2 - 2uv - 3v^2 \\ 3u^2 - 2uv - v^2 & -u^2 - 2uv + 3v^2 \end{pmatrix}_{\mathbf{u}=\mathbf{a}} \\ &= \begin{pmatrix} -5 & -7 \\ 3 & 15 \end{pmatrix} \end{aligned}$$

ii First calculate $\mathbf{b} = \mathbf{g}\left((1, -2)^T\right) = (-1, 3)^T$. Then

$$J\mathbf{f}(\mathbf{b}) = \begin{pmatrix} 2xy & x^2 \\ y^2 & 2xy \end{pmatrix}_{\mathbf{x}=\mathbf{b}} = \begin{pmatrix} -6 & 1 \\ 9 & -6 \end{pmatrix}.$$

Also

$$J\mathbf{g}(\mathbf{u}) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

for all $\mathbf{u}$, and in particular $\mathbf{u} = \mathbf{a}$.

The Chain Rules states, in terms of Jacobian matrices, that $J(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = \mathbf{f}\big(\mathbf{g}(\mathbf{a})\big) J\mathbf{g}(\mathbf{a})$. Here

$$J\mathbf{f}\big(\mathbf{g}(\mathbf{a})\big) J\mathbf{g}(\mathbf{a}) = \begin{pmatrix} -6 & 1 \\ 9 & -6 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} -5 & -7 \\ 3 & 15 \end{pmatrix}.$$

The same answer as in part i.!

7. Use the Chain Rule to find the Fréchet derivative of $\mathbf{f} \circ \mathbf{g}$ at the given point $\mathbf{a}$ for each of the following.

i. i. With $\mathbf{x} = (x, y)^T$, $\mathbf{u} = (u, v)^T \in \mathbb{R}^2$,

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} x^2 y \\ x - y \end{pmatrix} \quad \text{and} \quad \mathbf{g}(\mathbf{u}) = \begin{pmatrix} 3uv \\ u^2 - v \end{pmatrix},$$

at $\mathbf{a} = (2, 1)^T$.

ii. ii. With $\mathbf{x} = (x, y, z)^T \in \mathbb{R}^3$, $\mathbf{u} = (u, v)^T \in \mathbb{R}^2$,

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} xy \\ yz \end{pmatrix} \quad \text{and} \quad \mathbf{g}(\mathbf{u}) = \begin{pmatrix} uv^2 - v \\ u^2 \\ 1/uv \end{pmatrix},$$

at $\mathbf{a} = (2, 1)^T$.

Solution The $\mathbf{g}$ in part ii is not Fréchet differentiable at $\mathbf{0}$, but otherwise all functions are differentiable at all points in which we are interested. Thus, by the Composition Rule, $\mathbf{f} \circ \mathbf{g}$ is differentiable. Therefore $d(\mathbf{f} \circ \mathbf{g})_{\mathbf{a}}(\mathbf{t}) = J(\mathbf{f} \circ \mathbf{g})(\mathbf{a})(\mathbf{t})$ for all $\mathbf{t} \in \mathbb{R}^2$. The Chain Rule for matrices states that $J(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = J\mathbf{f}(\mathbf{b})\, J\mathbf{g}(\mathbf{a})$ where $\mathbf{b} = \mathbf{g}(\mathbf{a})$ and it is this product we will calculate.

i. First, $\mathbf{b} = \mathbf{g}(\mathbf{a}) = (6, 3)^T$. Then

$$J\mathbf{g}(\mathbf{a}) = \begin{pmatrix} 3v & 3u \\ 2u & -1 \end{pmatrix}_{\mathbf{u}=\mathbf{a}} = \begin{pmatrix} 3 & 6 \\ 4 & -1 \end{pmatrix}.$$

And

$$J\mathbf{f}(\mathbf{b}) = \begin{pmatrix} 2xy & x^2 \\ 1 & -1 \end{pmatrix}_{\mathbf{x}=\mathbf{b}} = \begin{pmatrix} 36 & 36 \\ 1 & -1 \end{pmatrix}.$$

Thus

$$(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = \begin{pmatrix} 36 & 36 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 3 & 6 \\ 4 & -1 \end{pmatrix} = \begin{pmatrix} 252 & 180 \\ -1 & 7 \end{pmatrix}.$$

The question asked you to find the Fréchet derivative which is, with $\mathbf{t} = (s,t)^T \in \mathbb{R}^2$,

$$d(\mathbf{f}\circ\mathbf{g})_{\mathbf{a}}(\mathbf{t}) = \begin{pmatrix} 252 & 180 \\ -1 & 7 \end{pmatrix}\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 252s + 180t \\ -s + 7t \end{pmatrix}.$$

ii. First, $\mathbf{b} = \mathbf{g}(\mathbf{a}) = (1,\, 4,\, 1/2)^T$. Then

$$J\mathbf{g}(\mathbf{a}) = \left.\begin{pmatrix} v^2 & 2uv - 1 \\ 2u & 0 \\ -1/u^2v & -1/uv^2 \end{pmatrix}\right|_{\mathbf{u}=\mathbf{a}} = \begin{pmatrix} 1 & 3 \\ 4 & 0 \\ -1/4 & -1/2 \end{pmatrix}$$

And

$$J\mathbf{f}(\mathbf{x}) = \left.\begin{pmatrix} y & x & 0 \\ 0 & z & y \end{pmatrix}\right|_{\mathbf{x}=\mathbf{b}} = \begin{pmatrix} 4 & 1 & 0 \\ 0 & 1/2 & 4 \end{pmatrix}.$$

Thus

$$(\mathbf{f}\circ\mathbf{g})(\mathbf{a}) = \begin{pmatrix} 4 & 1 & 0 \\ 0 & 1/2 & 4 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 4 & 0 \\ -1/4 & -1/2 \end{pmatrix} = \begin{pmatrix} 8 & 12 \\ 1 & -2 \end{pmatrix}.$$

The question asked you to find the Fréchet derivative which is, with $\mathbf{t} = (s,t)^T \in \mathbb{R}^2$,

$$d(\mathbf{f}\circ\mathbf{g})_{\mathbf{a}}(\mathbf{t}) = \begin{pmatrix} 8 & 12 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 8s + 12t \\ s - 2t \end{pmatrix}.$$

8. Consider the Chain Rule in the case

$$\mathbb{R}^p \xrightarrow{\mathbf{g}} \mathbb{R}^n \xrightarrow{f} \mathbb{R},$$

so f is scalar-valued. Assume $\mathbf{g}$ is Fréchet differentiable at $\mathbf{a} \in \mathbb{R}^p$ and f is Fréchet differentiable at $\mathbf{b} = \mathbf{g}(\mathbf{a}) \in \mathbb{R}^m$. The Chain Rule says that $f \circ \mathbf{g}$ is Fréchet differentiable at $\mathbf{a}$ and $J(f\circ\mathbf{g})(\mathbf{a}) = Jf(\mathbf{b})\, J\mathbf{g}(\mathbf{a})$.

Think of the coordinates in $\mathbb{R}^p$ as x^i for $1 \le i \le p$, while in $\mathbb{R}^n$ they will be y^j for $1 \le j \le n$. Show that the Chain Rule can be written as

$$\frac{\partial f\circ\mathbf{g}}{\partial x^i}(\mathbf{a}) = \sum_{k=1}^{n} \frac{\partial f}{\partial y^k}(\mathbf{b})\,\frac{\partial g^k}{\partial x^i}(\mathbf{a}),$$

for $1 \le i \le p$.

Solution Since f and $f \circ \mathbf{g}$ are scalar-valued functions their Jacobian matrices consist of only one row. In particular

$$Jf(\mathbf{b}) = (d_1 f(\mathbf{b}), ..., d_n f(\mathbf{b})) = \left(\frac{\partial f}{\partial y^1}(\mathbf{b}), ..., \frac{\partial f}{\partial y^n}(\mathbf{b}) \right),$$

Similarly

$$\begin{aligned} J(f \circ \mathbf{g})(\mathbf{a}) &= (d_1 (f \circ \mathbf{g})(\mathbf{a}), ..., d_p (f \circ \mathbf{g})(\mathbf{a})) \\ &= \left(\frac{\partial (f \circ \mathbf{g})}{\partial x^1}(\mathbf{a}), ..., \frac{\partial (f \circ \mathbf{g})}{\partial x^p}(\mathbf{a}) \right). \end{aligned}$$

From the definition of matrix multiplication the Chain Rule $J(f \circ \mathbf{g})(\mathbf{a}) = Jf(\mathbf{b})\, J\mathbf{g}(\mathbf{a})$ can be reinterpreted as saying that $d_i (f \circ \mathbf{g})(\mathbf{a})$, the i-th coordinate of $J(f \circ \mathbf{g})(\mathbf{a})$, equals the matrix product of $Jf(\mathbf{b})$ with the i-th column of $J\mathbf{g}(\mathbf{a})$, which is $d_i \mathbf{g}(\mathbf{a})$. This can be written in a number of ways.

First, for $1 \le i \le p$,

$$d_i (f \circ \mathbf{g})(\mathbf{a}) = Jf(\mathbf{b})\, d_i \mathbf{g}(\mathbf{a}) = \sum_{k=1}^{n} d_k f(\mathbf{b}) \left(d_i \mathbf{g}(\mathbf{a})\right)^k = \sum_{k=1}^{n} d_k f(\mathbf{b})\, d_i g^k(\mathbf{a}).$$

Or, with the alternative way of writing the partial derivatives,

$$\frac{\partial f \circ \mathbf{g}}{\partial x^i}(\mathbf{a}) = \sum_{k=1}^{n} \frac{\partial f}{\partial y^k}(\mathbf{b}) \frac{\partial g^k}{\partial x^i}(\mathbf{a}).$$

Extremal values of $d_{\mathbf{v}} f(\mathbf{a})$.

Here we find $\max_{\mathbf{v}:|\mathbf{v}|=1} d_{\mathbf{v}} f(\mathbf{a})$ *and* $\min_{\mathbf{v}:|\mathbf{v}|=1} d_{\mathbf{v}} f(\mathbf{a})$, *that is the directions of maximum and mimimum rate of change of* f *as we move away from* $\mathbf{a}$.

9. Suppose that $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is Fréchet differentiable on U and $\mathbf{a} \in U$. Prove that the directional derivative $d_{\mathbf{v}} f(\mathbf{a})$ has a

i. maximum value of $|\nabla f(\mathbf{a})|$ when $\mathbf{v}$ is in the direction of $\nabla f(\mathbf{a})$ and

ii. a minimum value of $-|\nabla f(\mathbf{a})|$ when $\mathbf{v}$ is in the direction of $-\nabla f(\mathbf{a})$.

Hint for any vectors we have $\mathbf{a} \bullet \mathbf{b} = |\mathbf{a}| \, |\mathbf{b}| \cos\theta$ where θ is the angle between the vectors $\mathbf{a}$ and $\mathbf{b}$.

Solution Since f is Fréchet differentiable on U and $\mathbf{a} \in U$ we have from the notes that for a unit vector $\mathbf{v}$, $d_{\mathbf{v}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \bullet \mathbf{v}$. Thus, by the hint in the question,

$$d_{\mathbf{v}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \bullet \mathbf{v} = |\nabla f(\mathbf{a})| \, |\mathbf{v}| \cos\theta = |\nabla f(\mathbf{a})| \cos\theta,$$

since $|\mathbf{v}| = 1$. Therefore, since $-1 \leq \cos\theta \leq 1$,

$$-|\nabla f(\mathbf{a})| \leq d_{\mathbf{v}} f(\mathbf{a}) \leq |\nabla f(\mathbf{a})| \,. \tag{3}$$

i. The upper bound is attained when the angle between $\mathbf{v}$ and $\nabla f(\mathbf{a})$ is 0, i.e. when $\mathbf{v}$ is in the direction of $\nabla f(\mathbf{a})$, which, since $\mathbf{v}$ is a unit vector, is when $\mathbf{v} = \nabla f(\mathbf{a}) / |\nabla f(\mathbf{a})|$. Then, for this value of $\mathbf{v}$,

$$d_{\mathbf{v}} f(\mathbf{a}) = \nabla f(\mathbf{a}) \bullet \frac{\nabla f(\mathbf{a})}{|\nabla f(\mathbf{a})|} = \frac{|\nabla f(\mathbf{a})|^2}{|\nabla f(\mathbf{a})|} = |\nabla f(\mathbf{a})| \,.$$

ii The lower bound in (3) is attained when the angle between $\mathbf{v}$ and $\nabla f(\mathbf{a})$ is π, i.e. when $\mathbf{v}$ is in the direction of $-\nabla f(\mathbf{a})$, which, since $\mathbf{v}$ is a unit vector, is when $\mathbf{v} = -\nabla f(\mathbf{a}) / |\nabla f(\mathbf{a})|$. Then, for this value of $\mathbf{v}$,

$$d_{\mathbf{v}} f(\mathbf{a}) = -\nabla f(\mathbf{a}) \bullet \frac{\nabla f(\mathbf{a})}{|\nabla f(\mathbf{a})|} = -\frac{|\nabla f(\mathbf{a})|^2}{|\nabla f(\mathbf{a})|} = -|\nabla f(\mathbf{a})| \,.$$

10. Suppose the temperature at a point $(x, y, z)^T$ in a metal cube is given by

$$T = 80 - 60xe^{-\frac{1}{20}\left(x^2+y^2+z^2\right)},$$

where the centre of the cube is taken to be $(0,0,0)^T$. In which direction from the origin is the rate of change of temperature greatest? The least?

Solution For simplicity let $r(\mathbf{x}) = (x^2 + y^2 + z^2)/20$. The gradient of T is

$$\nabla T(\mathbf{x}) = \begin{pmatrix} -60e^{-r(\mathbf{x})} + 6x^2 e^{-r(\mathbf{x})} \\ 6xye^{-r(\mathbf{x})} \\ 6xze^{-r(\mathbf{x})} \end{pmatrix} \text{ so } \nabla T(\mathbf{0}) = \begin{pmatrix} -60 \\ 0 \\ 0 \end{pmatrix}.$$

Hence the greatest rate of change is in the x-axis direction, $(-1,0,0)^T$, the least in the $(1,0,0)^T$ direction.

Solutions to Additional Questions 5

11 Define the function $f : \mathbb{R}^3 \to \mathbb{R}$ by $\mathbf{x} \mapsto xy^2z$.

i. Show that f is a C^1-function on $\mathbb{R}^3$.

ii. Calculate $\nabla f(\mathbf{a}) \bullet \mathbf{v}$ with $\mathbf{a} = (1, 3, -2)^T$ and $\mathbf{v} = (-1, 1, -2)^T / \sqrt{6}$. Explain any similarity with Question 4 Sheet 3.

Solution i. The gradient vector is

$$\nabla f(\mathbf{x}) = \begin{pmatrix} y^2z \\ 2xyz \\ xy^2 \end{pmatrix}.$$

ii.

$$\nabla f(\mathbf{a}) \bullet \mathbf{v} = \frac{1}{\sqrt{6}} \begin{pmatrix} -18 \\ -12 \\ 9 \end{pmatrix} \bullet \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix} = -\frac{12}{\sqrt{6}}.$$

This is the same as $d_{\mathbf{v}}f(\mathbf{a})$ which you were asked to calculate in Question 4 on Sheet 3. They are the same because f is, by part i., a C^1-function and thus Fréchet differentiable at $\mathbf{a}$. This is necessary to justify $d_{\mathbf{v}}f(\mathbf{a}) = \nabla f(\mathbf{a}) \bullet \mathbf{v}$.

12. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(\mathbf{x}) = \frac{\sin(x^2y^2)}{x^2 + y^2} \quad \text{if } \mathbf{x} = (x, y)^T \neq \mathbf{0}; \qquad f(\mathbf{0}) = 0.$$

i. Find the partial derivatives of f at **all** points $\mathbf{x} \in \mathbb{R}^2$.

   Hint For $\mathbf{x} = \mathbf{0}$ you will have to return to the definition of partial derivative.

ii. Prove that f is a C^1-function on $\mathbb{R}^2$ with Fréchet derivative $df_{\mathbf{0}} = \mathbf{0} : \mathbb{R}^2 \to \mathbb{R}$ at the origin.

   Hint You may make use of $|\sin\theta| \leq |\theta|$ for all θ.

Solution i. Partial differentiation gives

$$\frac{\partial f}{\partial x}(\mathbf{x}) = \frac{2xy^2\cos(x^2y^2)}{x^2 + y^2} - \frac{2x\sin(x^2y^2)}{(x^2 + y^2)^2},$$

$$\frac{\partial f}{\partial y}(\mathbf{x}) = \frac{2x^2y\cos(x^2y^2)}{x^2 + y^2} - \frac{2y\sin(x^2y^2)}{(x^2 + y^2)^2},$$

for $\mathbf{x} \neq \mathbf{0}$. For $\mathbf{x} = \mathbf{0}$ we return to the definition of differentiation,

$$\frac{\partial f}{\partial x}(\mathbf{0}) = \lim_{t\to 0} \frac{f(t\mathbf{e}_1)}{t} = \lim_{t\to 0} \frac{0}{t^3} = 0.$$

Similarly

$$\frac{\partial f}{\partial y}(\mathbf{0}) = \lim_{t\to 0} \frac{f(t\mathbf{e}_2)}{t} = \lim_{t\to 0} \frac{0}{t^3} = 0.$$

ii. The partial derivatives given in part i for $\mathbf{x} \neq \mathbf{0}$ are continuous wherever they are defined, i.e. $\mathbf{x} \neq \mathbf{0}$. For $\mathbf{x} = \mathbf{0}$ we have to return to the definition of continuity, that the limit equals the value of the function. Consider

$$\begin{aligned} \left|\frac{\partial f}{\partial x}(\mathbf{x}) - \frac{\partial f}{\partial x}(\mathbf{0})\right| &= \left|\frac{2xy^2\cos(x^2y^2)}{x^2+y^2} - \frac{2x\sin(x^2y^2)}{(x^2+y^2)^2} - 0\right| \\ &\leq \frac{2|x||y|^2}{|\mathbf{x}|^2} + \frac{2|x||x^2y^2|}{|\mathbf{x}|^4} \end{aligned}$$

using the triangle inequality along with $|\cos\theta| \leq 1$ and $|\sin\theta| \leq |\theta|$ for all θ. Then recalling that $|x|, |y| \leq |\mathbf{x}|$ we find that

$$\left|\frac{\partial f}{\partial x}(\mathbf{x}) - \frac{\partial f}{\partial x}(\mathbf{0})\right| \leq 4|\mathbf{x}| \to 0$$

as $\mathbf{x} \to 0$. Thus,

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{\partial f}{\partial x}(\mathbf{x}) = \frac{\partial f}{\partial x}(\mathbf{0}),$$

which is the definition that $\partial f(\mathbf{x})/\partial x$ is continuous at $\mathbf{x} = \mathbf{0}$. Similarly for $\partial f(\mathbf{x})/\partial y$. Hence f is a C^1-function at $\mathbf{x} = \mathbf{0}$ and thus on all of $\mathbb{R}^2$.

Therefore f is Fréchet differentiable on $\mathbb{R}^2$ and in particular at $\mathbf{0}$. This implies that

$$df_{\mathbf{0}}(\mathbf{v}) = \nabla f(\mathbf{0}) \bullet \mathbf{v} = \begin{pmatrix} \partial f(\mathbf{0})/\partial x \\ \partial f(\mathbf{0})/\partial y \end{pmatrix} \bullet \mathbf{v} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \bullet \mathbf{v} = \mathbf{0}.$$

True for all unit vectors $\mathbf{v}$ means that $df_{\mathbf{0}} = \mathbf{0} : \mathbb{R}^3 \to \mathbb{R}$. (That is, $df_{\mathbf{0}}$ is the linear map which send all vectors from $\mathbb{R}^3$ to 0 in $\mathbb{R}$.)

13. *Further practice on the Chain Rule* Use the chain rule to find the derivative of $\mathbf{f} \circ \mathbf{g}$ at the point $\mathbf{c}$ for each of the following. Give your answers in the form $d(\mathbf{f} \circ \mathbf{g})_{\mathbf{c}}(\mathbf{t})$.

i. $\mathbf{f}\big((x,y)^T\big) = (x^2y,\ x-y)^T$, $\mathbf{g}\big((u,v)^T\big) = (3uv,\ u^2-4v)^T$, $\mathbf{c} = (1,-2)^T$,

ii. $\mathbf{f}\big((x,y,z)^T\big) = (4xy,\ 3xz)^T$, $\mathbf{g}\big((u,v)^T\big) = \left(uv^2-4v,\ u^2,\ 4/uv\right)^T$, $\mathbf{c} = (-2,3)^T$

iii. $\mathbf{f}\big((x,y)^T\big) = (3x+4y,\ 2x^2y,\ x-y)^T$, $\mathbf{g}\big((u,v,w)^T\big) = (4u-3v+w,\ uv^2)^T$, $\mathbf{c} = (1,-2,3)^T$.

Solution The Chain Rule states that $d\left(\mathbf{f}\circ\mathbf{g}\right)_{\mathbf{c}} = d\mathbf{f}_{\mathbf{g}(\mathbf{c})}\circ d\mathbf{g}_{\mathbf{c}}$, so

$$d\left(\mathbf{f}\circ\mathbf{g}\right)_{\mathbf{c}}(\mathbf{t}) = d\mathbf{f}_{\mathbf{g}(\mathbf{c})}\left(d\mathbf{g}_{\mathbf{c}}(\mathbf{t})\right).$$

i. For $\mathbf{x} = (x,y)^T$, $\mathbf{u} = (u,v)^T$ and $\mathbf{t} = (s,t)^T \in \mathbb{R}^2$ we have

$$d\mathbf{f}_{\mathbf{x}}(\mathbf{u}) = Jf(\mathbf{x})\,\mathbf{u} = \begin{pmatrix} 2xy & x^2 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}.$$

and

$$d\mathbf{g}_{\mathbf{c}}(\mathbf{t}) = \begin{pmatrix} 3v & 3u \\ 2u & -4 \end{pmatrix}_{(1,-2)^T}\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} -6 & 3 \\ 2 & -4 \end{pmatrix}\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} -6s+3t \\ 2s-4t \end{pmatrix}.$$

Next $\mathbf{g}(\mathbf{c}) = (-6,9)^T$ so

$$\begin{aligned} d\mathbf{f}_{\mathbf{g}(\mathbf{c})}\left(d\mathbf{g}_{\mathbf{c}}(\mathbf{t})\right) &= \begin{pmatrix} 2xy & x^2 \\ 1 & -1 \end{pmatrix}_{(-6,9)^T}\begin{pmatrix} -6s+3t \\ 2s-4t \end{pmatrix} = \begin{pmatrix} -108 & 36 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} -6s+3t \\ 2s-4t \end{pmatrix} \\ &= \begin{pmatrix} 720s-468t \\ -8s+7t \end{pmatrix}. \end{aligned}$$

ii. For $\mathbf{x} = (x,y,z)^T$, $\mathbf{u} = (u,v)^T$ and $\mathbf{t} = (s,t)^T \in \mathbb{R}^2$ we have

$$d\mathbf{f}_{\mathbf{x}}(\mathbf{u}) = \begin{pmatrix} 4y & 4x & 0 \\ 3z & 0 & 3x \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}.$$

and

$$d\mathbf{g}_{\mathbf{c}}(\mathbf{t}) = \begin{pmatrix} v^2 & 2uv-4 \\ 2u & 0 \\ -4/u^2v & -4/uv^2 \end{pmatrix}_{(-2,3)^T}\mathbf{t} = \begin{pmatrix} 9 & -16 \\ -4 & 0 \\ -1/3 & 2/9 \end{pmatrix}\mathbf{t} = \begin{pmatrix} 9s-16t \\ -4s \\ -s/3+2t/9 \end{pmatrix}.$$

Next $\mathbf{g}\left((-2,3)^T\right) = (-30,\, 4,\, -2/3)^T$ so

$$\begin{aligned}
\mathbf{f}_{\mathbf{g}(\mathbf{c})}\left(d\mathbf{g}_{\mathbf{c}}(\mathbf{t})\right) &= \begin{pmatrix} 4y & 4x & 0 \\ 3z & 0 & 3x \end{pmatrix}_{\mathbf{x}=(-30,4,-2/3)^T} d\mathbf{g}_{\mathbf{c}}(\mathbf{t}) \\
&= \begin{pmatrix} 16 & -120 & 0 \\ -2 & 0 & -90 \end{pmatrix} \begin{pmatrix} 9s-16t \\ -4s \\ -s/3-2t/9 \end{pmatrix} \\
&= \begin{pmatrix} 624s - 256t \\ 12s + 42t \end{pmatrix}.
\end{aligned}$$

iii. An alternative approach is to not mention $\mathbf{t}$ until the end but, instead, look at the Jacobian matrices.

$$J\mathbf{g}(\mathbf{c}) = \begin{pmatrix} 4 & -3 & 1 \\ v^2 & 2uv & 0 \end{pmatrix}_{\mathbf{u}=\mathbf{c}} = \begin{pmatrix} 4 & -3 & 1 \\ 4 & -4 & 0 \end{pmatrix}.$$

Next $\mathbf{g}(\mathbf{c}) = (13,4)^T$. Then

$$J\mathbf{f}(\mathbf{g}(\mathbf{c})) = \begin{pmatrix} 3 & 4 \\ 4xy & 2x^2 \\ 1 & -1 \end{pmatrix}_{\mathbf{x}=(13,4)^T} = \begin{pmatrix} 3 & 4 \\ 208 & 338 \\ 1 & -1 \end{pmatrix}.$$

Multiplying together,

$$J\mathbf{f}(\mathbf{g}(\mathbf{c}))\, J\mathbf{g}(\mathbf{c}) = \begin{pmatrix} 3 & 4 \\ 208 & 338 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 4 & -3 & 1 \\ 4 & -4 & 0 \end{pmatrix} = \begin{pmatrix} 28 & -25 & 3 \\ 2184 & -1976 & 208 \\ 0 & 1 & 1 \end{pmatrix}.$$

Now introduce $\mathbf{t} \in \mathbb{R}^3$ so $\mathbf{t} = (r,s,t)^T$ say. Then

$$d\left(\mathbf{f}\circ\mathbf{g}\right)_{\mathbf{c}}(\mathbf{t}) = \begin{pmatrix} 28 & -25 & 3 \\ 2184 & -1976 & 208 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} r \\ s \\ t \end{pmatrix} = \begin{pmatrix} 28r - 25s + 3t \\ 2184r - 1976s + 208t \\ s+t \end{pmatrix}$$

14. Revisit Question 17iii on Sheet 3. Define the functions $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$ by $(x,y)^T \mapsto (x+y, x-y, xy)^T$ and $h : \mathbb{R}^3 \to \mathbb{R}$ by $(x,y,z)^T \mapsto xy^2z$.

Calculate, using the Chain Rule, the directional derivative of $h \circ \mathbf{f} : \mathbb{R}^2 \to \mathbb{R}$ at $\mathbf{a} = (2, -1)^T$ in the direction $\mathbf{v} = (1, -2)^T / \sqrt{5}$.

Solution Since $h \circ \mathbf{f}$ is scalar-valued we normally write $d_{\mathbf{v}} (h \circ \mathbf{f}) (\mathbf{a}) = \nabla (h \circ \mathbf{f}) (\mathbf{a}) \bullet \mathbf{v}$. But $\nabla (h \circ \mathbf{f}) (\mathbf{a}) = J (h \circ \mathbf{f}) (\mathbf{a})^T$ so $d_{\mathbf{v}} (h \circ \mathbf{f}) (\mathbf{a}) = J (h \circ \mathbf{f}) (\mathbf{a}) \mathbf{v}$. The Chain Rule states that

$$J (h \circ \mathbf{f}) (\mathbf{a}) = Jh\big(\mathbf{f}(\mathbf{a})\big) J\mathbf{f}(\mathbf{a}) = Jh(\mathbf{b})\, J\mathbf{f}(\mathbf{a})$$

with $\mathbf{b} = \mathbf{f}(\mathbf{a})$. In this case $\mathbf{b} = (1, 3, -2)^T$. The Jacobian matrices are

$$J\mathbf{f}(\mathbf{a}) = \left(\begin{array}{rr} 1 & 1 \\ 1 & -1 \\ y & x \end{array} \right)_{\mathbf{x}=\mathbf{a}} = \left(\begin{array}{rr} 1 & 1 \\ 1 & -1 \\ -1 & 2 \end{array} \right),$$

and

$$Jh(\mathbf{b}) = \left(\begin{array}{ccc} y^2 z & 2xyz & xy^2 \end{array} \right)_{\mathbf{x}=\mathbf{b}} = \left(\begin{array}{ccc} -18 & -12 & 9 \end{array} \right).$$

Then

$$J (h \circ \mathbf{f}) (\mathbf{a}) = \left(\begin{array}{ccc} -18 & -12 & 9 \end{array} \right) \left(\begin{array}{rr} 1 & 1 \\ 1 & -1 \\ -1 & 2 \end{array} \right) = \left(\begin{array}{cc} -39 & 12 \end{array} \right).$$

Finally,

$$\begin{aligned} d_{\mathbf{v}} (h \circ \mathbf{f}) (\mathbf{a}) &= \nabla (h \circ \mathbf{f}) (\mathbf{a}) \bullet \mathbf{v} \\ &= \frac{1}{\sqrt{5}} \left(\begin{array}{cc} -39 & 12 \end{array} \right) \left(\begin{array}{r} 1 \\ -2 \end{array} \right) = -\frac{63}{\sqrt{5}}. \end{aligned}$$

This should agree with your answer to Question 17 on Sheet 3. Would you agree that the calculations are simpler using the Chain Rule?

15. Assume $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^3$ is Fréchet differentiable at $\mathbf{q} = (2, 3)^T$ with

$$J\mathbf{F}(\mathbf{q}) = \left(\begin{array}{rr} -1 & 2 \\ 2 & -3 \\ 0 & 4 \end{array} \right).$$

Assume also that $\mathbf{F}(\mathbf{q}) = \left(\begin{array}{ccc} 2 & -1 & 4 \end{array} \right)^T$.

Define $f : \mathbb{R}^2 \to \mathbb{R} : f(\mathbf{x}) = |\mathbf{F}(x)|$. Prove that f is Fréchet differentiable at $\mathbf{q}$ and find $df_{\mathbf{q}} (\mathbf{t})$ for $\mathbf{t} \in \mathbb{R}^2$.

Solution The function f is the composition of $\mathbf{F}$ and the distance function $d(\mathbf{y}) := |\mathbf{y}|$ for $\mathbf{y} \in \mathbb{R}^3$. By assumption $\mathbf{F}$ is Fréchet differentiable at $\mathbf{q}$, and $|...|$ is Fréchet differentiable everywhere, in particular at $\mathbf{F}(\mathbf{q})$. Thus by the Chain rule f is Fréchet differentiable at $\mathbf{q}$.

It is simpler to first calculate the Jacobian matrix

$$Jf(\mathbf{q}) = J\left(d \circ \mathbf{F}\right)(\mathbf{q}) = Jd(\mathbf{F}(\mathbf{q}))\, J\mathbf{F}(\mathbf{q})\,.$$

For $\mathbf{y} = (x, y, z)^T \in \mathbb{R}^3, \mathbf{y} \neq \mathbf{0}$, we have $d(\mathbf{y}) = \left(x^2 + y^2 + z^2\right)^{1/2}$ so

$$Jd(\mathbf{y}) = \frac{1}{d(\mathbf{y})} \begin{pmatrix} x & y & z \end{pmatrix} = \frac{1}{d(\mathbf{y})} \mathbf{y}^T.$$

Thus, by the assumptions in the question,

$$\begin{aligned} Jf(\mathbf{q}) &= Jd(\mathbf{F}(\mathbf{q}))\, J\mathbf{F}(\mathbf{q}) = \frac{1}{d(\mathbf{F}(\mathbf{q}))} \mathbf{F}(\mathbf{q})^T \begin{pmatrix} -1 & 2 \\ 2 & -3 \\ 0 & 4 \end{pmatrix} \\ &= \frac{1}{\sqrt{21}} \begin{pmatrix} 2 & -1 & 4 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 2 & -3 \\ 0 & 4 \end{pmatrix} \\ &= \frac{1}{\sqrt{21}} \begin{pmatrix} 0 & 23 \end{pmatrix}. \end{aligned}$$

Finally, for $\mathbf{t} = (s, t)^T \in \mathbb{R}^2$ we have

$$df_{\mathbf{q}}(\mathbf{t}) = 23t/\sqrt{21}.$$

16. A heat-seeking insect always moves in the direction of the greatest increase in temperature. Describe the path of a heat-seeking insect placed at $(1, 1)^T$ on a metal plate heated so that the temperature at $\mathbf{x} = (x, y)^T$ is given by

$$T(\mathbf{x}) = 100 - 40xye^{-r(\mathbf{x})},$$

where $r(\mathbf{x}) = \left(x^2 + y^2\right)/10$.

What if the insect starts at $(3, 2)^T$? Or the origin $\mathbf{0}$?

Solution The gradient vector at $\mathbf{x} \in \mathbb{R}^2$ is

$$\nabla T(\mathbf{x}) = \begin{pmatrix} -40ye^{-r(\mathbf{x})} + 8x^2ye^{-r(\mathbf{x})} \\ -40xe^{-r(\mathbf{x})} + 8xy^2e^{-r(\mathbf{x})} \end{pmatrix} = e^{-r(\mathbf{x})} 8 \begin{pmatrix} -5y + x^2y \\ -5x + xy^2 \end{pmatrix}.$$

At time t the insect is at point $(x(t), y(t))^T$. It will be moving in the direction $(x'(t), y'(t))^T$. Being heat-seeking it will move in the direction of the greatest increase in temperature, given by $\nabla T(\mathbf{x})$. Thus $(x'(t), y'(t))^T = c\nabla T(\mathbf{x})$ for some $c > 0$. Therefore the ratio of coordinates are equal, i.e.

$$\frac{x'(t)}{y'(t)} = \frac{-5y + x^2 y}{-5x + xy^2} = \frac{y(x^2 - 5)}{x(y^2 - 5)},$$

as long as $y^2 \neq 5$. Rearrange as

$$\frac{x\, x'(t)}{5 - x^2} = \frac{y\, y'(t)}{5 - y^2}. \tag{4}$$

I have written it like this for at time 0 we are told $x = 1$ and so $5 - x^2 > 0$. The same also hold for $5 - y^2$. Integrate to get

$$-\frac{1}{2}\ln\left(5 - x^2\right) = -\frac{1}{2}\ln\left(5 - y^2\right) + C,$$

for a constant C, or

$$5 - x^2 = A\left(5 - y^2\right),$$

for $y^2 \neq 5$. where $A = e^{2C}$. To find A plug in the starting point $(1,1)^T$ to get $4 = 4A$, so $A = 1$. Thus the path is $x^2 = y^2$. or $x = \pm y$. The point $(1,1)^T$ does not lie on the line $x = -y$ so the answer is $x = y$. To find the direction of the line along which the insect travels look again at the gradient vector which, at $\mathbf{a} = (1,1)^T$ is, $\nabla T(\mathbf{a}) = e^{-r(\mathbf{a})} 8(-4,-4)^T$. This points towards the origin. Therefore, starting at $(1,1)^T$, the insect moves directly to the origin.

If the starting point is $(3,2)^T$ then you might have a reservation in using (4) for $5 - 3^2 < 0$ and so $\ln(5 - x^2)$ may well not be defined. You could, instead, write (4) as

$$-\frac{xx'(t)}{x^2 - 5} = \frac{yy'(t)}{5 - y^2}.$$

Integrate to get

$$-\frac{1}{2}\ln\left(x^2 - 5\right) = -\frac{1}{2}\ln\left(5 - y^2\right) + C,$$

or

$$x^2 - 5 = A\left(5 - y^2\right).$$

Plugging in the starting point $(3,2)^T$ we find $A = 4$ in which case

$$x^2 + 4y^2 = 25.$$

So at $(3,2)^T$ the insect starts on the path of this ellipse in the clockwise direction. (For the direction look at the signs of the components of $\nabla T\Big((3,2)^T\Big)$).

Be careful, this does not mean the insect traverses this ellipse without end - how could it do so gaining temperature all the time? This ellipse has been derived on the basis that $x^2 > 5$ and $y^2 < 5$. If either of these fails we have to re-examine the problem. One such point on the ellipse is $\mathbf{a} = \left(\sqrt{5}, -\sqrt{5}\right)^T$. But $\nabla T(\mathbf{a}) = \mathbf{0}$ so, at this point, the insect will not know which way to go and presumably stop.

At the origin the gradient vector $\nabla T(\mathbf{0})$ is $\mathbf{0}$ so again the insect will not know which way to go and will remain in place.