Solutions to Review Problems for Exam 1

1. Compute the (shortest) distance from the point $P(4, 0, -7)$ in $\mathbb{R}^3$ to the plane given by
$$4x - y - 3z = 12.$$

***Solution**:* The point $P_o(3, 0, 0)$ is in the plane. Let
$$w = \overrightarrow{P_oP} = \begin{pmatrix} 1 \\ 0 \\ -7 \end{pmatrix}$$

The vector $n = \begin{pmatrix} 4 \\ -1 \\ -3 \end{pmatrix}$ is orthogonal to the plane. To find the shortest distance, d, from P to the plane, we compute the norm of the orthogonal projection of w onto n; that is,
$$d = \|\text{Proj}_{\widehat{n}}(w)\|,$$
where
$$\widehat{n} = \frac{1}{\sqrt{26}} \begin{pmatrix} 4 \\ -1 \\ -3 \end{pmatrix},$$
a unit vector in the direction of n, and
$$\text{Proj}_{\widehat{n}}(w) = (w \cdot \widehat{n})\widehat{n}.$$
It then follows that
$$d = |w \cdot \widehat{n}|,$$
where $w \cdot \widehat{n} = \dfrac{1}{\sqrt{26}}(4 + 21) = \dfrac{25}{\sqrt{26}}$. Hence, $d = \dfrac{25\sqrt{26}}{26} \approx 4.9.$ □

2. Compute the (shortest) distance from the point $P(4, 0, -7)$ in $\mathbb{R}^3$ to the line given by the parametric equations
$$\begin{cases} x = -1 + 4t, \\ y = -7t, \\ z = 2 - t. \end{cases}$$

***Solution**:* The point $P_o(-1, 0, 2)$ is on the line. The vector

$$v = \begin{pmatrix} 4 \\ -7 \\ -1 \end{pmatrix}$$

gives the direction of the line. Put

$$w = \overrightarrow{P_oP} = \begin{pmatrix} 5 \\ 0 \\ -9 \end{pmatrix}.$$

The vectors v and w determine a parallelogram whose area is the norm of v times the shortest distance, d, from P to the line determined by v at P_o. We then have that

$$\text{area}(P(v, w)) = \|v\|d,$$

from which we get that

$$d = \frac{\text{area}(P(v, w))}{\|v\|}.$$

On the other hand,

$$\text{area}(P(v, w)) = \|v \times w\|,$$

where

$$v \times w = \begin{vmatrix} \widehat{i} & \widehat{j} & \widehat{k} \\ 4 & -7 & -1 \\ 5 & 0 & -9 \end{vmatrix} = 63\widehat{i} + 31\widehat{j} - 35\widehat{k}.$$

Thus, $\|v \times w\| = \sqrt{(63)^2 + (31)^2 + (35)^2} = \sqrt{6155}$ and therefore

$$d = \frac{\sqrt{6155}}{\sqrt{66}} \approx 9.7.$$

□

3. Compute the area of the triangle whose vertices in $\mathbb{R}^3$ are the points $(1, 1, 0)$, $(2, 0, 1)$ and $(0, 3, 1)$

***Solution**:* Label the points $P_o(1,1,0)$, $P_1(2,0,1)$ and $P_2(0,3,1)$ and define the vectors

$$v = \overrightarrow{P_oP_1} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \quad \text{and} \quad w = \overrightarrow{P_oP_2} = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}.$$

The area of the triangle determined by the points P_o, P_1 and P_2 is then half of the area of the parallelogram determined by the vectors v and w. Thus,

$$\text{area}(\triangle P_oP_1P_2) = \frac{1}{2}\|v \times w\|,$$

where

$$v \times w = \begin{vmatrix} \widehat{i} & \widehat{j} & \widehat{k} \\ 1 & -1 & 1 \\ -1 & 2 & 1 \end{vmatrix} = -3\widehat{i} - 2\widehat{j} + \widehat{k}.$$

Consequently, $\text{area}(\triangle P_oP_1P_2) = \dfrac{1}{2}\sqrt{9+4+1} = \dfrac{\sqrt{14}}{2} \approx 1.87.$ □

4. Let v and w be two vectors in $\mathbb{R}^3$, and let λ be a scalar. Show that the area of the parallelogram determined by the vectors v and $w + \lambda v$ is the same as that determined by v and w.

***Solution**:* The area of the parallelogram determined by v and $w+\lambda v$ is

$$\text{area}(P(v, w+\lambda v)) = \|v \times (w + \lambda v)\|,$$

where

$$v \times (w + \lambda v) = v \times w + \lambda v \times v = v \times w.$$

Consequently, $\text{area}(P(v, w+\lambda v)) = \|v \times w\| = \text{area}(P(v,w)).$ □

5. Let $\widehat{u}$ denote a unit vector in $\mathbb{R}^n$ and $P_{\widehat{u}}(v)$ denote the orthogonal projection of v along the direction of $\widehat{u}$ for any vector $v \in \mathbb{R}^n$. Use the Cauchy–Schwarz inequality to prove that the map

$$v \mapsto P_{\widehat{u}}(v) \quad \text{for all} \quad v \in \mathbb{R}^n$$

is a continuous map from $\mathbb{R}^n$ to $\mathbb{R}^n$.

Solution: $P_{\widehat{u}}(v) = (v \cdot \widehat{u})\widehat{u}$ for all $v \in \mathbb{R}^n$. Consequently, for any $w, v \in \mathbb{R}^n$,

$$\begin{aligned} P_{\widehat{u}}(w) - P_{\widehat{u}}(v) &= (w \cdot \widehat{u})\widehat{u} - (v \cdot \widehat{u})\widehat{u} \\ &= (w \cdot \widehat{u} - v \cdot \widehat{u})\widehat{u} \\ &= [(w - v) \cdot \widehat{u}]\widehat{u}. \end{aligned}$$

It then follows that

$$\|P_{\widehat{u}}(w) - P_{\widehat{u}}(v)\| = |(w - v) \cdot \widehat{u}|,$$

since $\|\widehat{u}\| = 1$. Hence, by the Cauchy–Schwarz inequality,

$$\|P_{\widehat{u}}(w) - P_{\widehat{u}}(v)\| \leqslant \|w - v\|.$$

Applying the Squeeze Theorem we then get that

$$\lim_{\|w-v\|\to 0} \|P_{\widehat{u}}(w) - P_{\widehat{u}}(v)\| = 0,$$

which shows that $P_{\widehat{u}}$ is continuous at every $v \in V$. □

6. Define the scalar field $f\colon \mathbb{R}^n \to \mathbb{R}$ by $f(v) = \dfrac{1}{2}\|v\|^2$ for all $v \in \mathbb{R}^n$. Show that f is differentiable on $\mathbb{R}^n$ and compute the linear map $Df(u)\colon \mathbb{R}^n \to \mathbb{R}$ for all $u \in \mathbb{R}^n$. What is the gradient of f at u for all $x \in \mathbb{R}^n$?

Solution: Let u and w be any vector in $\mathbb{R}^n$ and consider

$$\begin{aligned} f(u + w) &= \frac{1}{2}\|u + w\|^2 \\ &= \frac{1}{2}(u + w) \cdot (u + w) \\ &= \frac{1}{2}u \cdot u + u \cdot w + \frac{1}{2}w \cdot w \\ &= \frac{1}{2}\|u\|^2 + u \cdot w + \frac{1}{2}\|w\|^2. \end{aligned}$$

Thus,

$$f(u + w) - f(u) - u \cdot w = \frac{1}{2}\|w\|^2.$$

Consequently,

$$\frac{|f(u+w) - f(u) - u\cdot w|}{\|w\|} = \frac{1}{2}\|w\|,$$

from which we get that

$$\lim_{\|w\|\to 0} \frac{|f(u+w) - f(u) - u\cdot w|}{\|w\|} = 0,$$

and therefore f is differentiable at u with derivative map $Df(u)$ given by

$$Df(u)w = u\cdot w \quad \text{for all } w\in\mathbb{R}^n.$$

Hence, $\nabla f(u) = u$ for all $u\in\mathbb{R}^n$. □

7. Let $g\colon [0,\infty)\to\mathbb{R}$ be a differentiable, real–valued function of a single variable, and let $f(x,y) = g(r)$ where $r = \sqrt{x^2+y^2}$.

   (a) Compute $\dfrac{\partial r}{\partial x}$ in terms of x and r, and $\dfrac{\partial r}{\partial y}$ in terms of y and r.

   Solution: Take the partial derivative of $r^2 = x^2 + y^2$ on both sides with respect to x to obtain

   $$\frac{\partial(r^2)}{\partial x} = 2x.$$

   Applying the chain rule on the left–hand side we get

   $$2r\frac{\partial r}{\partial x} = 2x,$$

   which leads to

   $$\frac{\partial r}{\partial x} = \frac{x}{r}.$$

   Similarly, $\dfrac{\partial r}{\partial y} = \dfrac{y}{r}$. □

   (b) Compute ∇f in terms of $g'(r)$, r and the vector $\mathbf{r} = x\hat{i} + y\hat{j}$.

   Solution: Take the partial derivative of $f(x,y) = g(r)$ on both sides with respect to x and apply the Chain Rule to obtain

   $$\frac{\partial f}{\partial x} = g'(r)\frac{\partial r}{\partial x} = g'(r)\frac{x}{r}.$$

Similarly, $\dfrac{\partial f}{\partial y} = g'(r)\dfrac{y}{r}$.
It then follows that

$$\begin{aligned} \nabla f &= \frac{\partial f}{\partial x}\widehat{i} + \frac{\partial f}{\partial y}\widehat{j} \\ &= g'(r)\frac{x}{r}\widehat{i} + g'(r)\frac{y}{r}\widehat{j} \\ &= \frac{g'(r)}{r}(x\widehat{i} + y\widehat{j}) \\ &= \frac{g'(r)}{r}\mathbf{r}. \end{aligned}$$

□

8. Let $f\colon U \to \mathbb{R}$ denote a scalar field defined on an open subset U of $\mathbb{R}^n$, and let $\widehat{u}$ be a unit vector in $\mathbb{R}^n$. If the limit

$$\lim_{t\to 0} \frac{f(v + t\widehat{u}) - f(v)}{t}$$

exists, we call it the *directional derivative of f at v in the direction of the unit vector $\widehat{u}$*. We denote it by $D_{\widehat{u}}f(v)$.

(a) Show that if f is differentiable at $v \in U$, then, for any unit vector $\widehat{u}$ in $\mathbb{R}^n$, the directional derivative of f in the direction of $\widehat{u}$ at v exists, and

$$D_{\widehat{u}}f(v) = \nabla f(v) \cdot \widehat{u},$$

where $\nabla f(v)$ is the gradient of f at v.

Proof: Suppose that f is differentiable at $v \in U$. Then,

$$f(v + w) = f(v) + Df(v)w + E(w),$$

where

$$Df(v)w = \nabla f(v) \cdot w,$$

and

$$\lim_{\|w\|\to 0} \frac{|E(w)|}{\|w\|} = 0.$$

Thus, for any $t \in \mathbb{R}$,

$$f(v + t\widehat{u}) = f(v) + t\nabla f(v) \cdot \widehat{u} + E(t\widehat{u}),$$

where

$$\lim_{|t| \to 0} \frac{|E(t\widehat{u})|}{|t|} = 0,$$

since $\|t\widehat{u}\| = |t|\|\widehat{u}\| = |t|$.

We then have that, for $t \neq 0$,

$$\frac{f(v + t\widehat{u}) - f(v)}{t} - \nabla f(v) \cdot \widehat{u} = \frac{E(t\widehat{u})}{t},$$

and consequently

$$\left| \frac{f(v + t\widehat{u}) - f(v)}{t} - \nabla f(v) \cdot \widehat{u} \right| = \frac{|E(t\widehat{u})|}{|t|},$$

from which we get that

$$\lim_{t \to 0} \left| \frac{f(v + t\widehat{u}) - f(v)}{t} - \nabla f(v) \cdot \widehat{u} \right| = 0.$$

$\square$

(b) Suppose that $f \colon U \to \mathbb{R}$ is differentiable at $v \in U$. Prove that if $D_{\widehat{u}} f(v) = 0$ for every unit vector $\widehat{u}$ in $\mathbb{R}^n$, then $\nabla f(v)$ must be the zero vector.

Proof: Suppose, by way of contradiction, that $\nabla f(v) \neq \mathbf{0}$, and put

$$\widehat{u} = \frac{1}{\|\nabla f(v)\|} \nabla f(v).$$

Then, $\widehat{u}$ is a unit vector, and therefore, by the assumption,

$$D_{\widehat{u}} f(v) = 0,$$

or

$$\nabla f(v) \cdot \widehat{u} = 0.$$

But this implies that

$$\nabla f(v) \cdot \frac{1}{\|\nabla f(v)\|} \nabla f(v) = 0,$$

where

$$\begin{aligned}\nabla f(v)\cdot\frac{1}{\|\nabla f(v)\|}\nabla f(v) &= \frac{1}{\|\nabla f(v)\|}\nabla f(v)\cdot\nabla f(v)\\ &= \frac{1}{\|\nabla f(v)\|}\|\nabla f(v)\|^2\\ &= \|\nabla f(v)\|.\end{aligned}$$

It then follows that $\|\nabla f(v)\| = 0$, which contradicts the assumption that $\nabla f(v) \neq \mathbf{0}$. Therefore, $\nabla f(v)$ must be the zero vector. □

(c) Suppose that $f\colon U \to \mathbb{R}$ is differentiable at $v \in U$. Use the Cauchy–Schwarz inequality to show that the largest value of $D_{\widehat{u}}f(v)$ is $\|\nabla f(v)\|$ and it occurs when $\widehat{u}$ is in the direction of $\nabla f(v)$.

Proof. If f is differentiable at x, then $D_{\widehat{u}}f(x) = \nabla f(x)\cdot\widehat{u}$, as was shown in part (a). Thus, by the Cauchy–Schwarz inequality,

$$|D_{\widehat{u}}f(x)| \leqslant \|\nabla f(x)\|\|\widehat{u}\| = \|\nabla f(x)\|,$$

since $\widehat{u}$ is a unit vector. Hence,

$$-\|\nabla f(x)\| \leqslant D_{\widehat{u}}f(x) \leqslant \|\nabla f(x)\|$$

for any unit vector $\widehat{u}$, and so the largest value that $D_{\widehat{u}}f(x)$ can have is $\|\nabla f(x)\|$.

If $\nabla f(x) \neq \mathbf{0}$, then $\widehat{u} = \dfrac{1}{\|\nabla f(x)\|}\nabla f(x)$ is a unit vector, and

$$\begin{aligned}D_{\widehat{u}}f(x) &= \nabla f(x)\cdot\widehat{u}\\ &= \nabla f(x)\cdot\frac{1}{\|\nabla f(x)\|}\nabla f(x)\\ &= \frac{1}{\|\nabla f(x)\|}\nabla f(x)\cdot\nabla f(x)\\ &= \frac{1}{\|\nabla f(x)\|}\|\nabla f(x)\|^2\\ &= \|\nabla f(x)\|.\end{aligned}$$

Thus, $D_{\widehat{u}}f(x)$ attains its largest value when $\widehat{u}$ is in the direction of $\nabla f(x)$. □

9. The scalar field $f\colon U \to \mathbb{R}$ is said to have a *local minimum* at $x \in U$ if there exists $r > 0$ such that $B_r(x) \subseteq U$ and

$$f(x) \leqslant f(y) \quad \text{for every} \;\; y \in B_r(x).$$

Prove that if f is differentiable at $x \in U$ and f has a local minimum at x, then $\nabla f(x) = \mathbf{0}$, the zero vector in $\mathbb{R}^n$.

Proof. Let $\widehat{u}$ be a unit vector and $t \in \mathbb{R}$ be such that $|t| < r$; then,

$$f(x + t\widehat{u}) \geqslant f(x),$$

from which we get that

$$f(x + t\widehat{u}) - f(x) \geqslant 0.$$

Dividing by $t > 0$ we then have that

$$\frac{f(x + t\widehat{u}) - f(x)}{t} \geqslant 0.$$

Thus, letting $t \to 0^+$, we get that

$$D_{\widehat{u}} f(x) \geqslant 0, \tag{1}$$

since f is differentiable at x. Similarly, dividing by $t < 0$, we have

$$\frac{f(x + t\widehat{u}) - f(x)}{t} \leqslant 0,$$

from which we obtain, letting $t \to 0^-$, that

$$D_{\widehat{u}} f(x) \leqslant 0. \tag{2}$$

Combining (1) and (2) we then have that

$$D_{\widehat{u}} f(x) = 0,$$

where $\widehat{u}$ is an arbitrary unit vector. It then follows from the previous problem that $\nabla f(x) = \mathbf{0}$. □

10. Let I denote an open interval in $\mathbb{R}$. Suppose that $\sigma\colon I \to \mathbb{R}^n$ and $\gamma\colon I \to \mathbb{R}^n$ are paths in $\mathbb{R}^n$. Define a real valued function $f\colon I \to \mathbb{R}$ of a single variable by

$$f(t) = \sigma(t) \cdot \gamma(t) \quad \text{for all} \;\; t \in I;$$

that is, $f(t)$ is the dot product of the two paths at t.

Show that if σ and γ are both differentiable on I, then so is f, and

$$f'(t) = \sigma'(t) \cdot \gamma(t) + \sigma(t) \cdot \gamma'(t) \quad \text{for all} \;\; t \in I.$$

***Solution**:* Let $t \in I$ and assume that both σ and γ are differentiable at t. Then,

$$\sigma(t+h) = \sigma(t) + h\sigma'(t) + E_1(h), \quad \text{for } |h| \text{ sufficiently small,}$$

where

$$\lim_{h\to 0} \frac{\|E_1(h)\|}{|h|} = 0. \tag{3}$$

Similarly,

$$\gamma(t+h) = \gamma(t) + h\gamma'(t) + E_2(h), \quad \text{for } |h| \text{ sufficiently small,}$$

where

$$\lim_{h\to 0} \frac{\|E_2(h)\|}{|h|} = 0. \tag{4}$$

It then follows that, for $|h|$ sufficiently small,

$$\begin{aligned}
f(t+h) &= \sigma(t+h)\cdot\gamma(t+h) \\
&= (\sigma(t) + h\sigma'(t) + E_1(h))\cdot(\gamma(t) + h\gamma'(t) + E_2(h)) \\
&= \sigma(t)\cdot\gamma(t) + h\sigma(t)\cdot\gamma'(t) + \sigma(t)\cdot E_2(h)) + h\sigma'(t)\cdot\gamma(t) \\
&\quad +h^2\sigma'(t)\cdot\gamma'(t) + h\sigma'(t)\cdot E_2(h) + E_1(h)\cdot\gamma(t) \\
&\quad +hE_1(h)\cdot\gamma'(t) + E_1(h)\cdot E_2(h) \\
&= f(t) + h[\sigma(t)\cdot\gamma'(t) + \sigma'(t)\cdot\gamma(t)] + h^2\sigma'(t)\cdot\gamma'(t) \\
&\quad +\sigma(t)\cdot E_2(h)) + h\sigma'(t)\cdot E_2(h) + E_1(h)\cdot\gamma(t) \\
&\quad +hE_1(h)\cdot\gamma'(t) + E_1(h)\cdot E_2(h)
\end{aligned}$$

Rearranging terms and dividing by $h \neq 0$ and $|h|$ small enough, we then have that

$$\begin{aligned}
\frac{f(t+h)-f(t)}{h} &= \sigma(t)\cdot\gamma'(t) + \sigma'(t)\cdot\gamma(t) + +h\sigma'(t)\cdot\gamma'(t) \\
&\quad +\sigma(t)\cdot\frac{E_2(h)}{h} + \sigma'(t)\cdot E_2(h) + \frac{E_1(h)}{h}\cdot\gamma(t) \\
&\quad +E_1(h)\cdot\gamma'(t) + E_1(h)\cdot\frac{E_2(h)}{h}
\end{aligned}$$

Observe that, as $h \to 0$, all the terms on the right hand side of the previous expression which involve E_1 or E_2 go to 0, by virtue of the

Cauchy–Schwarz inequality and (3) and (4). Therefore, we obtain that

$$\lim_{h\to 0}\frac{f(t+h)-f(t)}{h} = \sigma(t)\cdot\gamma'(t) + \sigma'(t)\cdot\gamma(t).$$

Hence, f is differentiable at t, and its derivative at t is

$$f'(t) = \sigma(t)\cdot\gamma'(t) + \sigma'(t)\cdot\gamma(t).$$

Since t is an arbitrary element of I, the result follows. □

11. Let $\sigma\colon I \to \mathbb{R}^n$ denote a differentiable path in $\mathbb{R}^n$. Show that if $\|\sigma(t)\|$ is constant for all $t\in I$, then $\sigma'(t)$ is orthogonal to $\sigma(t)$ for all $t \in I$.

    ***Solution**:* Let $\|\sigma(t)\| = c$, where c denotes a constant. Then,

    $$\|\sigma(t)\|^2 = c^2,$$

    or

    $$\sigma(t)\cdot\sigma(t) = c^2.$$

    Differentiating with respect to t on both sides, and using the result of the previous problem, we obtain that

    $$\sigma(t)\cdot\sigma'(t) + \sigma'(t)\cdot\sigma(t) = 0,$$

    or, by the symmetry of the dot–product,

    $$2\sigma'(t)\cdot\sigma(t) = 0,$$

    or

    $$\sigma'(t)\cdot\sigma(t) = 0.$$

    Hence, $\sigma'(t)$ is orthogonal to $\sigma(t)$ for all $t\in I$. □

12. A particle is following a path in three–dimensional space given by

    $$\sigma(t) = (e^t, e^{-t}, 1-t) \quad \text{for} \quad t\in\mathbb{R}.$$

    At time $t_o = 1$, the particle flies off on a tangent.

    (a) Where will the particle be at time $t_1 = 2$?

Solution: Find the tangent line to the path at $\sigma(1)$:

$$\overrightarrow{r}(t) = \sigma(1) + (t-1)\sigma'(1),$$

where

$$\sigma'(t) = (e^t, -e^{-t}, -1) \quad \text{for} \quad t \in \mathbb{R}.$$

Then,

$$\overrightarrow{r}(t) = (e, 1/e, 0) + (t-1)(e, -1/e, -1).$$

The parametric equations of the tangent line then are

$$\begin{cases} x = e + e(t-1) \\ y = 1/e - (t-1)/e \\ z = 1 - t \end{cases}$$

When $t = 2$, the particle will be at the point in $\mathbb{R}^3$ with coordinates

$$(2e, 0, -1).$$

□

(b) Will the particle ever hit the xy–plane? Is so, find the location on the xy plane where the particle hits.

Answer: The particle leaves the path at the point with coordinates $(e, 1/e, 0)$ on the xy–plane. After that, it doesn't come back to it. □

13. Let U denote an open and convex subset of $\mathbb{R}^n$. Suppose that $f\colon U \to \mathbb{R}$ is differentiable at every $x \in U$. Fix x and y in U, and define $g\colon [0,1] \to \mathbb{R}$ by

$$g(t) = f(x + t(y-x)) \quad \text{for} \quad 0 \leqslant t \leqslant 1.$$

(a) Explain why the function g is well defined.

Solution: Since U is convex, $x + t(y-x)$ is in U for $0 \leqslant t \leqslant 1$. Thus, $f(x + t(y-x))$ is defined for $t \in [0,1]$. □

(b) Show that g is differentiable on $(0,1)$ and that

$$g'(t) = \nabla f(x + t(y-x)) \cdot (y-x) \quad \text{for} \quad 0 < t < 1.$$

(*Suggestion:* Consider

$$\frac{g(t+h) - g(t)}{h} = \frac{f(x + t(y-x) + h(y-x)) - f(x + t(y-x))}{h}$$

and apply the definition of differentiability of f at the point $x + t(y-x)$.)

Proof. Since f is differentiable on U, for $|h|$ small enough,

$$f(x+t(y-x)+h(y-x)) = f(x+t(y-x))+Df(x+t(y-x))(h(y-x))+E(h(y-x)),$$

where

$$\lim_{\|w\|\to 0} \frac{|E(w)|}{\|w\|} = 0. \tag{5}$$

Thus,

$$f(x+t(y-x)+h(y-x)) = f(x+t(y-x))+h\nabla f(x+t(y-x))\cdot(y-x)+E((h(y-x)),$$

from which we get that

$$\begin{aligned} \frac{g(t+h)-g(t)}{h} &= \frac{f(x+t(y-x)+h(y-x))-f(x+t(y-x))}{h} \\ &= \nabla f(x+t(y-x))\cdot(y-x)+\frac{E(h(y-x))}{h} \end{aligned}$$

for $h \neq 0$.
Observe that

$$\lim_{h\to 0} \frac{|E(h(y-x))|}{h} = \lim_{h\to 0} \|y-x\| \frac{|E(h(y-x))|}{\|h(y-x)\|} = 0,$$

by virtue of (5). It then follows that

$$\lim_{h\to 0} \frac{g(t+h)-g(t)}{h} = \nabla f(x+t(y-x))\cdot(y-x),$$

and therefore g is differentiable at t and $g'(t) = \nabla f(x+t(y-x))\cdot(y-x)$. □

(c) Use the Mean Value Theorem for derivatives to show that there exists a point z is the line segment connecting x to y such that

$$f(y)-f(x) = D_{\widehat{u}} f(z)\|y-x\|,$$

where $\widehat{u}$ is the unit vector in the direction of the vector $y-x$; that is, $\widehat{u} = \dfrac{1}{\|y-x\|}(y-x)$.
(*Hint:* Observe that $g(1)-g(0) = f(y)-f(x)$.)

Solution: Assume that $x \neq y$, for if $x = y$ the equality certainly holds true.

By the Mean Value Theorem, there exists $\tau \in (0,1)$ such that

$$g(1) - g(0) = g'(\tau)(1-0) = g'(\tau).$$

It then follows that

$$f(y) - f(x) = \nabla f(x + \tau(y-x)) \cdot (y-x).$$

Put $z = x + \tau(y-x)$; then, z is a point in the line segment connecting x to y, and

$$\begin{aligned} f(y) - f(x) &= \nabla f(z) \cdot (y-x) \\ &= \nabla f(z) \cdot \frac{y-x}{\|y-x\|}\|y-x\| \\ &= \nabla f(z) \cdot \widehat{u}\ \|y-x\| \\ &= D_{\widehat{u}} f(z)\|y-x\|, \end{aligned}$$

where $\widehat{u} = \dfrac{1}{\|y-x\|}(y-x)$. □

14. Prove that if U is an open and convex subset of $\mathbb{R}^n$, and $f\colon U \to \mathbb{R}$ is differentiable on U with $\nabla f(v) = \mathbf{0}$ for all $v \in U$, then f must be a constant function.

***Solution**:* Fix $x_o \in U$; then, since U is convex, for any $x \in U\backslash\{x_o\}$, the line segment connecting x_o to x is entirely contained in U. Furthermore, by the argument in part (c) of the previous problem, there exists z in the line segment connecting x_o to x such that

$$f(x) - f(x_o) = D_{\widehat{u}} f(z)\|x - x_o\|,$$

where $\widehat{u} = \dfrac{1}{\|x-x_o\|}(x - x_o)$.
Now, $D_{\widehat{u}} f(z) = \nabla f(z) \cdot \widehat{u} = 0$, since $\nabla f(x) = \mathbf{0}$ for all $x \in U$. Therefore,

$$f(x) = f(x_o).$$

Since x was arbitrary, it follows that f maps every element in U to $f(x_o)$; that is, f is a constant function. □