Derivative of a Composition

It may be necessary, for various reasons, to change variables in a given function. For example, one may change from Cartesian coordinates (x, y) to polar coordinates (r, θ) in a function like

$$f(x, y) = x^2 - y^2.$$

The polar and Cartesian coordinates are related by the equations

$$x = r\cos\theta \qquad \text{and} \qquad y = r\sin\theta$$

therefore such a change would give a new function

$$F(r, \theta) = r^2\left(\cos^2\theta - \sin^2\theta\right) = r^2\cos 2\theta$$

which may be more convenient to handle than f, (since it involves one term $r^2\cos 2\theta$). Note that $F(r, \theta)$ is really a composition of $f(x, y)$ with $g(r, \theta) = (r\cos\theta, r\sin\theta)$.

This section introduces a formula for the derivative of a composition $f \circ g$ in terms of the derivatives of f and g. It turns out to be a generalization of the expression for the derivative of a composition of functions of one variable. To revisit that formula, let f and g be functions of one variable. Suppose g is differentiable at a point c in its domain with derivative $g'(c)$. Furthermore, suppose f is differentiable at $g(c)$ with derivative $f'(g(c))$. (We are assuming that $g(c)$ is in the domain of f.) Then, by the chain rule, $f \circ g$ is differentiable at c with derivative

$$(f \circ g)'(c) = f'(g(c)) \cdot g'(c)$$

Thus the derivative of $f \circ g$ is obtained by multiplying the derivative of f at $g(c)$ with the derivative of g at c.

Now consider functions $g(u, v) : A \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ and $f(x, y) : B \subseteq \mathbb{R}^2 \to \mathbb{R}$. Assume that $g(A) \subset B$ so that the composition $F = f \circ g$ is defined. For convenience, assume that g is a function of variables u and v while f is a function of another set of variables x and y. Since g is from a subset of $\mathbb{R}^2$ into $\mathbb{R}^2$, it has two components:

$$g(u, v) = (g_1(u, v), g_2(u, v))$$

Suppose g is differentiable at a point $(c, d) \in A$. Its Jacobian matrix at (c, d) is

$$\begin{pmatrix} \dfrac{\partial g_1(c,d)}{\partial u} & \dfrac{\partial g_1(c,d)}{\partial v} \\ \dfrac{\partial g_2(c,d)}{\partial u} & \dfrac{\partial g_2(c,d)}{\partial v} \end{pmatrix}$$

Suppose, in addition, f is differentiable at $g(c, d)$. Its Jacobian matrix at $g(c, d)$ is

$$\begin{pmatrix} \dfrac{\partial f(g(c,d))}{\partial x} & \dfrac{\partial f(g(c,d))}{\partial y} \end{pmatrix}$$

According to the chain rule for functions of several variables, $F = f \circ g$ is differentiable at (c, d) and its Jacobian matrix $\begin{pmatrix} \dfrac{\partial F(c,d)}{\partial u} & \dfrac{\partial F(c,d)}{\partial v} \end{pmatrix}$ at (c, d) is given by

$$\begin{pmatrix} \dfrac{\partial F(c,d)}{\partial u} & \dfrac{\partial F(c,d)}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f(g(c,d))}{\partial x} & \dfrac{\partial f(g(c,d))}{\partial y} \end{pmatrix} \begin{pmatrix} \dfrac{\partial g_1(c,d)}{\partial u} & \dfrac{\partial g_1(c,d)}{\partial v} \\ \dfrac{\partial g_2(c,d)}{\partial u} & \dfrac{\partial g_2(c,d)}{\partial v} \end{pmatrix} \tag{1}$$

In other words, the Jacobian matrix for $F = f \circ g$ is the product of the two Jacobian matrices for f and g in that order. Since (1) is a product of matrices and matrix multiplication is not commutative, the order of the product is important.

Example 1 *Let* $f(x,y) = x^2 - y^2$ *and* $g(u,v) = (u\cos v, u\sin v)$. *Form the composition* $F(u,v) = f \circ g$. *Consider the point* $(2, \frac{\pi}{3})$ *in* $\mathbb{R}^2$. *The Jacobian matrix for* g *at* $(2, \frac{\pi}{3})$ *is*

$$\begin{pmatrix} \cos\frac{\pi}{3} & -2\sin\frac{\pi}{3} \\ \sin\frac{\pi}{3} & 2\cos\frac{\pi}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\sqrt{3} \\ \frac{\sqrt{3}}{2} & 1 \end{pmatrix}$$

and the Jacobian matrix for f *at* $(1, \sqrt{3})$, *(i.e. at* $g(2, \frac{\pi}{3})$*), is*

$$\begin{pmatrix} 2 & -2\sqrt{3} \end{pmatrix}$$

By the chain rule, the Jacobian matrix of $F = f \circ g$ *at* $(2, \frac{\pi}{3})$ *is*

$$\begin{pmatrix} 2 & -2\sqrt{3} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\sqrt{3} \\ \frac{\sqrt{3}}{2} & 1 \end{pmatrix} = \begin{pmatrix} -2 & -4\sqrt{3} \end{pmatrix}$$

This implies that $f \circ g(2+h, \frac{\pi}{3}+k) - f \circ g\ (2, \frac{\pi}{3}) \simeq -2h - 4\sqrt{3}k$ *and the error in this approximation shrinks to* 0 *faster than* $||(h,k)||$.

Since we have the formulas for f *and* g, *we may calculate the Jacobian matrix directly. Indeed*

$$F(u,v) = f \circ g\,(u,v) = f\,(u\cos v, u\sin v) = u^2\cos 2v$$

Therefore $\dfrac{\partial F(u,v)}{\partial u} = 2u\cos 2v$ *and* $\dfrac{\partial F(u,v)}{\partial v} = -2u^2\sin 2v$, *hence the Jacobian matrix of the composition at* $(2, \frac{\pi}{3})$ *is*

$$\begin{pmatrix} \dfrac{\partial F(2,\frac{\pi}{3})}{\partial u} & \dfrac{\partial F(2,\frac{\pi}{3})}{\partial v} \end{pmatrix} = \begin{pmatrix} 2(2)\cos\frac{2\pi}{3} & -2(2)^2\sin\frac{2\pi}{3} \end{pmatrix} = \begin{pmatrix} -2 & -4\sqrt{3} \end{pmatrix}$$

In general, the Jacobian matrix of $F = f \circ g$ *at an arbitrary point* (u,v) *is*

$$\begin{pmatrix} 2u\cos v & -2u\sin v \end{pmatrix} \begin{pmatrix} \cos v & -u\sin v \\ \sin v & u\cos v \end{pmatrix} = \begin{pmatrix} 2u\cos^2 v - 2u\sin^2 v & -4u\cos v\sin v \end{pmatrix}$$

$$= \begin{pmatrix} 2u\cos 2v & -2u\sin 2v \end{pmatrix}$$

Standard Notation

Let $g(u,v) : A \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ and $f(x,y) : B \subseteq \mathbb{R}^2 \to \mathbb{R}$ be given functions with $g(A) \subseteq B$, so that the composition $F = f \circ g$ is defined. Since g is a function from $\mathbb{R}^2$ into $\mathbb{R}^2$, it has two components, therefore we may write it as

$$g(u,v) = (g_1(u,v), g_2(u,v))$$

Using formula (1), the Jacobian matrix of $F = f \circ g$ at (u,v) is

$$\begin{pmatrix} \dfrac{\partial F\,(u,v)}{\partial u} & \dfrac{\partial F\,(u,v)}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f(g(u,v))}{\partial x} & \dfrac{\partial f(g(u,v))}{\partial y} \end{pmatrix} \begin{pmatrix} \dfrac{\partial g_1(u,v)}{\partial u} & \dfrac{\partial g_1(u,v)}{\partial v} \\ \dfrac{\partial g_2(u,v)}{\partial u} & \dfrac{\partial g_2(u,v)}{\partial v} \end{pmatrix}$$

Since $g_1(u,v)$ replaces x and $g_2(u,v)$ replaces y when we form the composition $f \circ g$, it is standard practice to write $x = g_1(u,v)$, $y = g_2(u,v)$ and $f(g(u,v)) = f(x,y)$. Consequently $\dfrac{\partial g_1(u,v)}{\partial u}$ is written as $\dfrac{\partial x}{\partial u}$ and

$\dfrac{\partial g_1(c,d)}{\partial v}$ as $\dfrac{\partial x}{\partial v}$. Similarly, $\dfrac{\partial g_2(u,v)}{\partial u}$ and $\dfrac{\partial g_2(u,v)}{\partial v}$ are written as $\dfrac{\partial y}{\partial u}$ and $\dfrac{\partial y}{\partial v}$ respectively. Consequently, the above matrix equation is written as

$$\begin{pmatrix} \dfrac{\partial F\,(u,v)}{\partial u} & \dfrac{\partial F\,(u,v)}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f(x,y)}{\partial x} & \dfrac{\partial f(x,y)}{\partial y} \end{pmatrix} \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix}$$

or simply

$$\begin{pmatrix} \dfrac{\partial F}{\partial u} & \dfrac{\partial F}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix}$$

The next pill may be hard to swallow; instead of regarding $f \circ g(u,v)$ as a new function $F(u,v)$, the practice is to write it as $f(u,v)$. Then its Jacobian matrix is written as $\begin{pmatrix} \dfrac{\partial f(u,v)}{\partial u} & \dfrac{\partial f(u,v)}{\partial v} \end{pmatrix}$ instead of $\begin{pmatrix} \dfrac{\partial F(u,v)}{\partial u} & \dfrac{\partial F(u,v)}{\partial v} \end{pmatrix}$. Thus, instead of writing

$$\begin{pmatrix} \dfrac{\partial F(u,v)}{\partial u} & \dfrac{\partial F(u,v)}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix}$$

it is common practice to write

$$\begin{pmatrix} \dfrac{\partial f(u,v)}{\partial u} & \dfrac{\partial f(u,v)}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix}$$

or simply

$$\begin{pmatrix} \dfrac{\partial f}{\partial u} & \dfrac{\partial f}{\partial v} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{pmatrix}$$

This is multiplied out to get

$$\begin{aligned} \frac{\partial f}{\partial u} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} \\ \frac{\partial f}{\partial v} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v} \end{aligned}$$

and it is generally called the chain rule for a function of two variables.

Example 2 *Let $f(x,y)$ be a given function and $g(r,\theta) = (r\cos\theta, r\sin\theta)$. Form the composition $F(r,\theta) = f(g(r,\theta))$. Write $x = r\cos\theta$ and $y = r\sin\theta$. Then $\dfrac{\partial F}{\partial r}$ and $\dfrac{\partial F}{\partial \theta}$ are given by*

$$\begin{aligned} \frac{\partial F}{\partial r} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} = \cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y} \\ \frac{\partial F}{\partial \theta} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = -r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y} \end{aligned}$$

Higher order partial derivatives are obtained by repeated differentiation. For example

$$\frac{\partial^2 F}{\partial r^2} = \frac{\partial}{\partial r}\left(\frac{\partial F}{\partial r}\right) = \frac{\partial}{\partial r}\left(\cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y}\right) = \cos\theta\frac{\partial}{\partial r}\left(\frac{\partial f}{\partial x}\right) + \sin\theta\frac{\partial}{\partial r}\left(\frac{\partial f}{\partial y}\right)$$

Remember that $\frac{\partial f}{\partial x}$ *is an abbreviation for* $\frac{\partial f(g(r,\theta))}{\partial x}$*, therefore* $\frac{\partial}{\partial r}\left(\frac{\partial f}{\partial x}\right)$ *is evaluated the same way we evaluated* $\frac{\partial}{\partial r}f(g(r,\theta))$*. The result is*

$$\frac{\partial}{\partial r}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\frac{\partial x}{\partial r} + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\frac{\partial y}{\partial r} = \left(\frac{\partial^2 f}{\partial x^2}\right)\cos\theta + \left(\frac{\partial^2 f}{\partial y\partial x}\right)\sin\theta$$

Similarly,

$$\frac{\partial}{\partial r}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\frac{\partial x}{\partial r} + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)\frac{\partial y}{\partial r} = \left(\frac{\partial^2 f}{\partial x\partial y}\right)\cos\theta + \left(\frac{\partial^2 f}{\partial y^2}\right)\sin\theta$$

Therefore

$$\begin{aligned}\frac{\partial^2 F}{\partial r^2} &= \cos\theta\left[\left(\frac{\partial^2 f}{\partial x^2}\right)\cos\theta + \left(\frac{\partial^2 f}{\partial y\partial x}\right)\sin\theta\right] + \sin\theta\left[\left(\frac{\partial^2 f}{\partial x\partial y}\right)\cos\theta + \left(\frac{\partial^2 f}{\partial y^2}\right)\sin\theta\right]\\ &= \left(\frac{\partial^2 f}{\partial x^2}\right)\cos^2\theta + 2\left(\frac{\partial^2 f}{\partial y\partial x}\right)\sin\theta\cos\theta + \left(\frac{\partial^2 f}{\partial y^2}\right)\sin^2\theta\end{aligned}$$

To determine $\frac{\partial^2 F}{\partial \theta^2}$*, one has to compute* $\frac{\partial}{\partial\theta}\left(\frac{\partial F}{\partial\theta}\right) = \frac{\partial}{\partial\theta}\left(-r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y}\right)$*. By the product rule*

$$\frac{\partial}{\partial\theta}\left(-r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y}\right) = -r\cos\theta\frac{\partial f}{\partial x} - r\sin\theta\frac{\partial}{\partial\theta}\left(\frac{\partial f}{\partial x}\right) - r\sin\theta\frac{\partial f}{\partial y} + r\cos\theta\frac{\partial}{\partial\theta}\left(\frac{\partial f}{\partial y}\right)$$

We now calculate the partial derivatives of $\frac{\partial f}{\partial x}$ *and* $\frac{\partial f}{\partial y}$ *with respect to* θ *the same way we calculated their partial derivatives with respect to* r*.*

$$\frac{\partial}{\partial\theta}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\frac{\partial x}{\partial\theta} + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\frac{\partial y}{\partial\theta} = \left(\frac{\partial^2 f}{\partial x^2}\right)(-r\sin\theta) + \left(\frac{\partial^2 f}{\partial y\partial x}\right)(r\cos\theta)$$

and

$$\frac{\partial}{\partial\theta}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\frac{\partial x}{\partial\theta} + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)\frac{\partial y}{\partial\theta} = \left(\frac{\partial^2 f}{\partial x\partial y}\right)(-r\sin\theta) + \left(\frac{\partial^2 f}{\partial y^2}\right)(r\cos\theta)$$

Therefore

$$\begin{aligned}\frac{\partial^2 F}{\partial\theta^2} &= -r\frac{\partial f}{\partial x}\cos\theta - r\sin\theta\left[\left(\frac{\partial^2 f}{\partial x^2}\right)(-r\sin\theta) + \left(\frac{\partial^2 f}{\partial y\partial x}\right)(r\cos\theta)\right] - r\frac{\partial f}{\partial y}\sin\theta\\ &\quad + r\cos\theta\left[\left(\frac{\partial^2 f}{\partial x\partial y}\right)(-r\sin\theta) + \left(\frac{\partial^2 f}{\partial y^2}\right)(r\cos\theta)\right]\end{aligned}$$

This may be reduced to

$$\frac{\partial^2 F}{\partial\theta^2} = -r\left[\frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta\right] + r^2\left(\frac{\partial^2 f}{\partial x^2}\right)\sin^2\theta - 2r^2\left(\frac{\partial^2 f}{\partial x\partial y}\right)\sin\theta\cos\theta + r^2\left(\frac{\partial^2 f}{\partial y^2}\right)\cos^2\theta$$

The chain rule extends to more general functions in the obvious way. Thus let $g : A \subseteq \mathbb{R}^p \to \mathbb{R}^q$ and $f : B \subseteq \mathbb{R}^q \to \mathbb{R}^m$ be given functions such that $g(A) \subseteq B$ so that the composition $f \circ g$ is defined. If g is differentiable at a point $c = (c_1, \ldots, c_p)$ in its domain g and f is differentiable at $g(c)$ then $f \circ g$ is differentiable at c and its derivative is related to the derivatives of f and g by the following identity:

$$D(f \circ g)(c)h = Df(g(c)) \circ Dg(c)h.$$

In particular:

Let the Jacobian matrix for g at c be

$$\begin{pmatrix} \dfrac{\partial g_1(c)}{\partial u_1} & \cdots & \dfrac{\partial g_1(c)}{\partial u_p} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial g_q(c)}{\partial u_1} & \cdots & \dfrac{\partial g_q(c)}{\partial u_p} \end{pmatrix}$$

Let the Jacobian matrix for f at $g(c)$ be

$$\begin{pmatrix} \dfrac{\partial f_1(g(c))}{\partial x_1} & \cdots & \dfrac{\partial f_1(g(c))}{\partial x_q} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial f_m(g(c))}{\partial x_1} & \cdots & \dfrac{\partial f_m(g(c))}{\partial x_q} \end{pmatrix}$$

Then the Jacobian matrix for $f \circ g$ at c is

$$\begin{pmatrix} \dfrac{\partial f_1(g(c))}{\partial x_1} & \cdots & \dfrac{\partial f_1(g(c))}{\partial x_q} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial f_m(g(c))}{\partial x_1} & \cdots & \dfrac{\partial f_m(g(c))}{\partial x_q} \end{pmatrix} \begin{pmatrix} \dfrac{\partial g_1(c)}{\partial u_1} & \cdots & \dfrac{\partial g_1(c)}{\partial u_p} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial g_q(c)}{\partial u_1} & \cdots & \dfrac{\partial g_q(c)}{\partial u_p} \end{pmatrix}$$

Therefore

$$D\,(f \circ g)\,(c)h = \begin{pmatrix} \dfrac{\partial f_1(g(c))}{\partial x_1} & \cdots & \dfrac{\partial f_1(g(c))}{\partial x_q} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial f_m(g(c))}{\partial x_1} & \cdots & \dfrac{\partial f_m(g(c))}{\partial x_q} \end{pmatrix} \begin{pmatrix} \dfrac{\partial g_1(c)}{\partial u_1} & \cdots & \dfrac{\partial g_1(c)}{\partial u_p} \\ \vdots & \vdots & \vdots \\ \dfrac{\partial g_q(c)}{\partial u_1} & \cdots & \dfrac{\partial g_q(c)}{\partial u_p} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_p \end{pmatrix}$$

Exercise 3 *Let $f : A \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a given function. Consider the composition $F = f \circ g$ for each given function $g : \mathbb{R}^2 \to \mathbb{R}^2$ then determine expressions for F_u, F_v, F_{uv}, F_{vv} and F_{uu}.*

1. $g(u, v) = (3u - 5v, 7u + 6v)$.
2. $g(u, v) = (-v + \ln u, u + \ln v)$.

Mean Value Theorem

Theorem 4 *Let $f(x, y)$ be a differentiable function from a subset of $\mathbb{R}^2$ into $\mathbb{R}$. Assume that the line segment ℓ joining (a_1, a_2) and (b_1, b_2) is in the domain of f. Denote $(b_1, b_2) - (a_1, a_2)$ by (h, k). Then there is a point (c_1, c_2) on ℓ such that*

$$f\,(b_1, b_2) - f\,(a_1, a_2) = Df\,(c_1, c_2)\,(h, k) = hf_x\,(c_1, c_2) + kf_y\,(c_1, c_2)$$

We prove the theorem by restricting f to points on ℓ. This essentially gives us a function of one variable to which we may apply the Mean Value theorem for a function of one variable. To this end, note that the points on ℓ have the form $(a_1, a_2) + t\,(h, k) = (a_1 + th, a_2 + tk)$ where $0 \leq t \leq 1$. Indeed, when $t = 0$ the expression $(a_1 + th, a_2 + tk)$ reduces to (a_1, a_2) and when $t = 1$ it simplifies to (b_1, b_2). The values of t between 0 and 1 generate points on ℓ in between the two end points. Now consider the function

$$g(t) = f(a_1 + th, a_2 + tk)$$

It is the composition of the two differentiable functions f and $u(t) = (a_1 + th, a_2 + tk)$. Further more, it is a function of one variable $\dot{t}$ that satisfies the two conditions

$$g(0) = f\,(a_1, a_2) \text{ and } g(1) = f\,(b_1, b_2)$$

By the mean value theorem for a function of one variable, there is a number θ between 0 and 1 such that

$$g(1) - g(0) = g'(\theta)\,(1 - 0)$$

Denote $u(\theta) = (a_1 + \theta h, a_2 + \theta k)$ by (c_1, c_2). Since $g(t) = f \circ u(t)$, the chain rule implies that

$$\begin{aligned} g'(\theta)\,(1-0) &= Df(c_1, c_2) \circ Du(\theta)\,(1-0) = \begin{pmatrix} f_x(c_1, c_2) & f_y(c_1, c_2) \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} \\ &= hf_x(c_1, c_2) + kf_y(c_1, c_2) \end{aligned}$$

The extension to functions of three or more variables is straight-forward. For example, if $f(x, y, z)$ is a differentiable function of three variables from a subset of $\mathbb{R}^3$ into $\mathbb{R}$ and the line segment ℓ joining points (a_1, a_2, a_3) and (b_1, b_2, b_3) is in the domain of f then there is a point (c_1, c_2, c_3) on ℓ such that

$$f\,(b_1, b_2, b_3) - f\,(a_1, a_2, a_3) = (b_1 - a_1) f_x\,(c_1, c_2, c_3) + (b_2 - a_2)\,f_y\,(c_1, c_2, c_3) + (b_3 - a_3)\,f_z\,(c_1, c_2, c_3)$$

As you will soon find out, the Mean Value Theorem is a powerful theoretical tool that is used in many situations.