Christopher Lum
lum@uw.edu

Lecture 04i
Chain Rule

The YouTube video entitled ‘The Chain Rule’ that covers this lecture is located at https://youtu.be/tf-pLFQB-7sU

Outline

-Introduction
-Chain Rule (Single Input Single Output)
-Chain Rule (Multiple Input Single Output)
    -Composite Function Formulation
    -Jacobian Matrix Formulation
-Chain Rule (Multiple Input Multiple Output)

Introduction

Example
Consider a composite function of

$$h(u) = \left(3u + 4u^2\right)^3$$

In[1]:=
```
h[u_] = (3 u + 4 u^2)^3;
```

To compute the derivative $\frac{dh}{du}$ we can simply take the derivative of the aforementioned function using “brute force” techniques (AKA take the derivative of the outside and then multiply by the derivative of the inner)

$$\frac{dh}{du} = \frac{d}{du}\left[\left(3u + 4u^2\right)^3\right]$$

$$= \left\{3\left(3u + 4u^2\right)^2\right\}\left\{\frac{d}{du}\left[3u + 4u^2\right]\right\}$$

$$= \{3(3u + 4u^2)^2\}\{3 + 8u\}$$

In[2]:=
```
dhdu[u_] = (3 * (3 u + 4 u^2)^2) × (3 + 8 u)
dhduCheck[u_] = D[h[u], u];
dhdu[u] == dhduCheck[u] // Simplify
```

Out[2]= $3 \times (3 + 8\,u)\,(3\,u + 4\,u^2)^2$

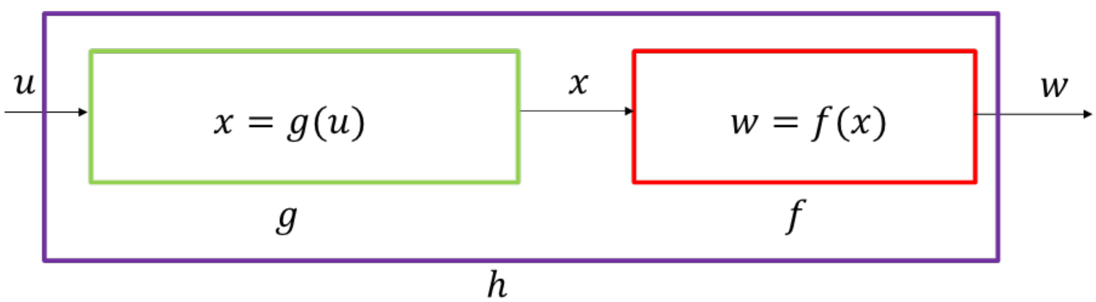
Out[4]= True

Chain Rule (Single Input Single Output)

The aforementioned brute force technique is actually an application of the Chain Rule which can be more easily understood if we consider h to be a composite function with a single input and single output

$$h(u) = f(g(u))$$

This can be visualized below



The standard Chain Rule you saw in high school calculus states that

$$\frac{dw}{du} = \frac{dw}{dx}\frac{dx}{du} \qquad \textbf{(Eq.1.a)}$$

Or more explicitly

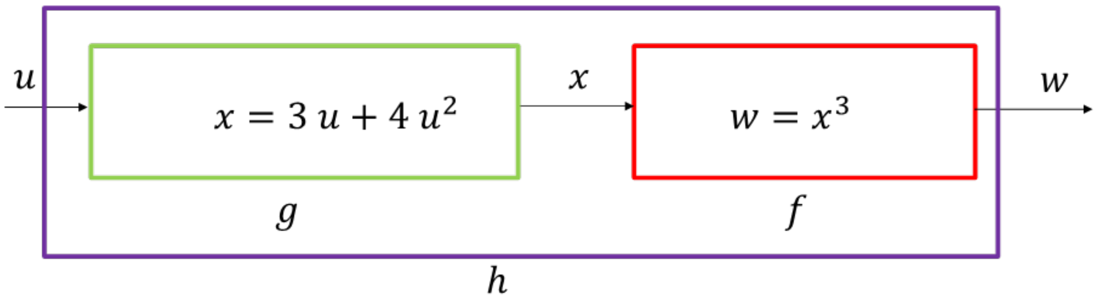
$$\frac{dw}{du}\Big|_u = \frac{dw}{dx}\Big|_{x(u)} \cdot \frac{dx}{du}\Big|_u \qquad \textbf{(Eq.1.b)}$$

We can alternatively use the Chain Rule on the aforementioned example by first formulating this as

$$x = g(u) = 3u + 4u^2$$
$$w = f(x) = x^3$$

In[5]:=
```
g[u_] = 3 u + 4 u^2;
f[x_] = x^3;
```

This is visualized below

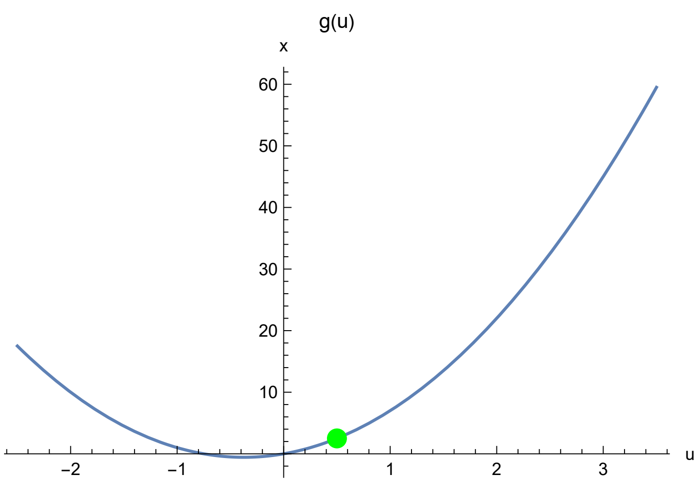


Conceptually, we can think of small changes in u affecting x. x is output of the g function, which is then the input to the f function so these changes in x will affect the output of the f function depending on where it is operating. Continuing our example consider $u = 1/2$

In[7]:=
```
uo = 1 / 2;
xo = g[uo]
xoP = Point[{uo, xo}];
Show[
 Plot[g[u], {u, uo - 3, uo + 3}],
 Graphics[{PointSize[0.03], Green, xoP}],
 AxesLabel → {"u", "x"},
 PlotLabel → "g(u)"
]
```

Out[8]= $\frac{5}{2}$

Out[10]=

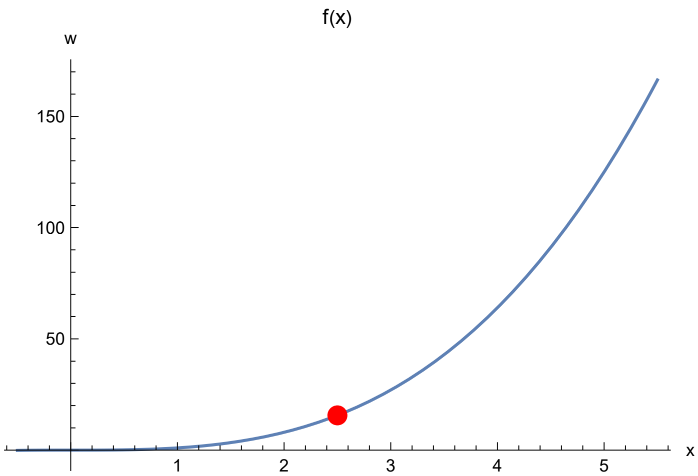


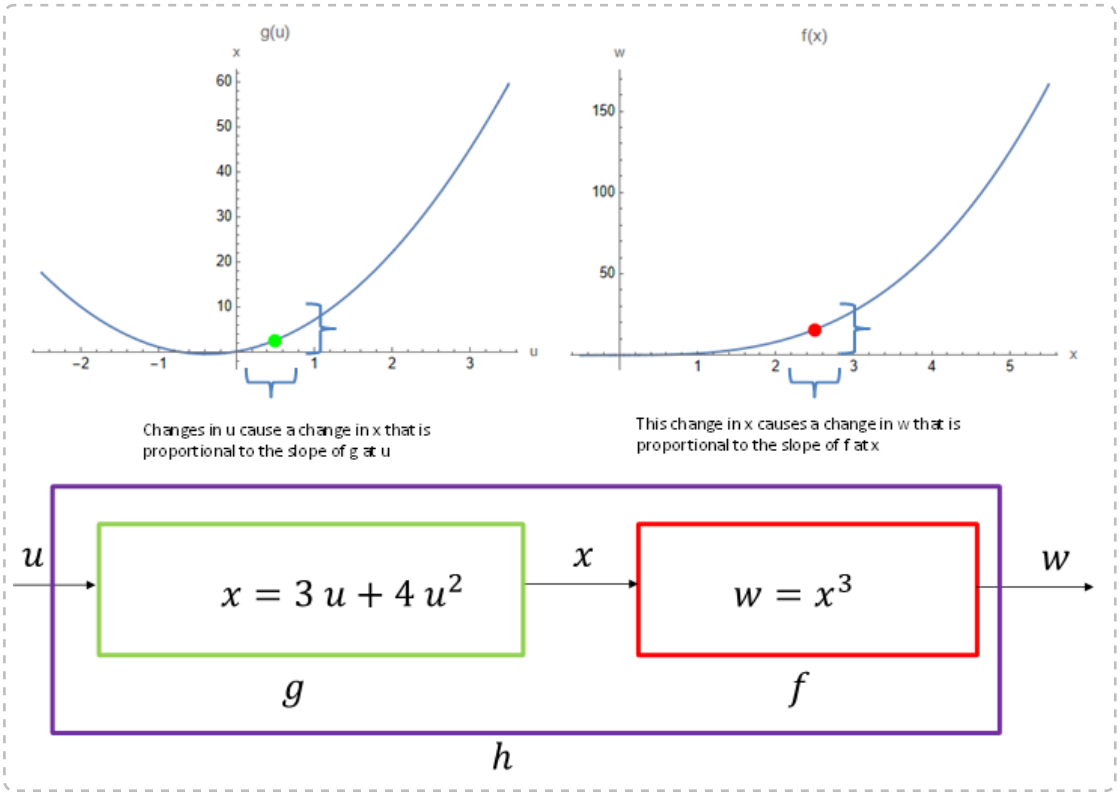
So the sensitivity of f at this point is

In[11]:=
```
wo = f[xo]
woP = Point[{xo, wo}];
Show[
 Plot[f[x], {x, xo - 3, xo + 3}],
 Graphics[{PointSize[0.03], Red, woP}],
 AxesLabel → {"x", "w"},
 PlotLabel → "f(x)"
]
```

Out[11]= $\frac{125}{8}$

Out[13]=





So by Eq.1.b we have

$$\frac{dw}{du}\,\Big|_{u} = \frac{dw}{dx}\,\Big|_{x(u)} \cdot \frac{dx}{du}\,\Big|_{u}$$

We can compute each element, starting with $\frac{dx}{du}\,\Big|_{u}$

$$\frac{dx}{du}\,\Big|_{u} = \frac{d}{du}\,\Big|_{u}\,[g(u)]$$

$$= \frac{d}{du}\,\Big|_{u}\left[3\,u + 4\,u^2\right]$$

$$= 3 + 8\,u$$

Also the term $\frac{dw}{dx}\,\Big|_{x(u)}$

$$\frac{dw}{dx}\,\Big|_{x(u)} = \frac{d}{dx}\,\Big|_{x(u)}\,[f(x)]$$

$$= \frac{d}{dx}\,\Big|_{x(u)}\left[x^3\right]$$

$$= 3\,x^2\,\Big|_{x(u)} \qquad \text{note: } x = 3\,u + 4\,u^2$$

$$= 3\left(3\,u + 4\,u^2\right)^2$$

So combining we have

$$\frac{dw}{du}\,|_u = 3\left(3u + 4u^2\right)^2 (3 + 8u)$$

In[•]:=
```
dwdu[u_] = 3 (3 u + 4 u^2)^2 (3 + 8 u)
```

Out[•]=
```
3 × (3 + 8 u) (3 u + 4 u^2)^2
```

This is the same as the expression we obtained before using “brute force” techniques

In[•]:=
```
dwdu[u] == dhdu[u]
```

Out[•]=
```
True
```

Chain Rule (Multiple Input Single Output)

Let us extend this idea to a function with multiple inputs and a single output.

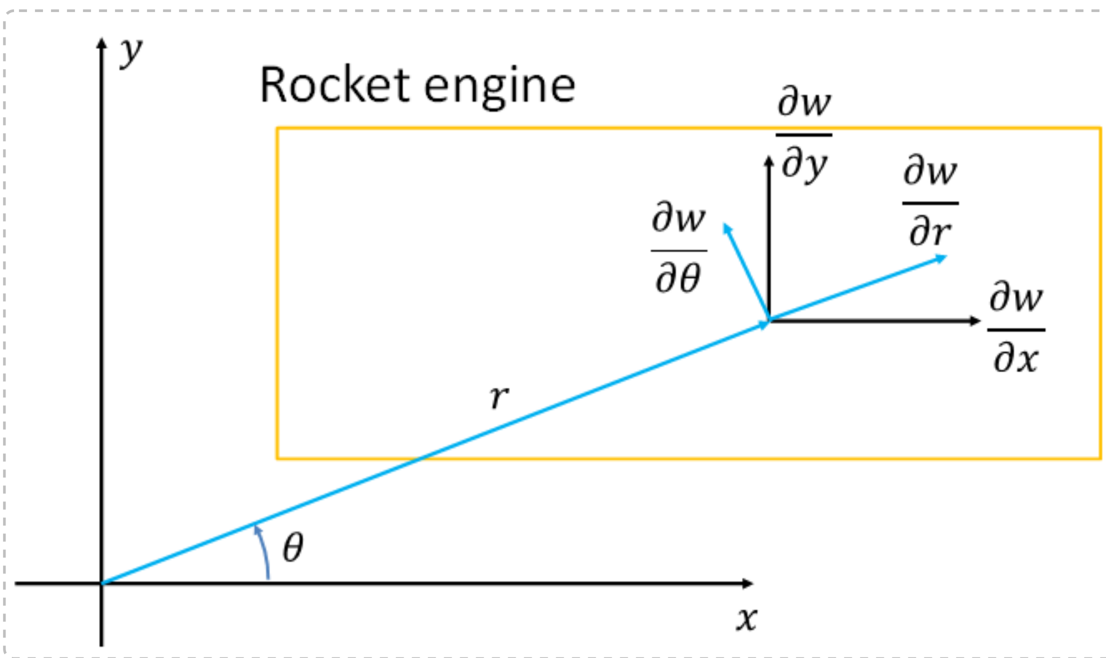
Example 1: Chain Rule

Consider the temperature inside the 3 dimensional combustion chamber of a rocket engine. The temperature at a point x, y, z is given as

$$w = 2x^2 y - 4y^2 + z^3$$

In[•]:=
```
w[x_, y_, z_] = 2 x^2 y - 4 y^2 + z^3;
```

A sample graphic is shown below



We can easily determine how the temperature changes in the x, y, and z directions by taking the partial derivatives

$$\frac{\partial w}{\partial x} = \frac{\partial}{\partial x}\left[2x^2 y - 4y^2 + z^3\right] = 4xy$$

$$\frac{\partial w}{\partial y} = \frac{\partial}{\partial y}\left[2x^2 y - 4y^2 + z^3\right] = 2x^2 - 8y$$

$$\frac{\partial w}{\partial z} = \frac{\partial}{\partial z}\left[2x^2 y - 4y^2 + z^3\right] = 3z^2$$

In[∘]:=
```
dwdx = D[w[x, y, z], x]
dwdy = D[w[x, y, z], y]
dwdz = D[w[x, y, z], z]
```

Out[∘]= $4\,x\,y$

Out[∘]= $2\,x^2 - 8\,y$

Out[∘]= $3\,z^2$

However, suppose we would like to describe the position using polar coordinates.

$$x = x(r, \theta, z) = r\cos(\theta)$$
$$y = y(r, \theta, z) = r\sin(\theta)$$
$$z = z(r, \theta, z) = z$$

In[∘]:=
```
x[r_, θ_, z_] = r Cos[θ];
y[r_, θ_, z_] = r Sin[θ];
z[r_, θ_, z_] = z;
```

So in this case, we see that our function w can be described as

$$w = f(x(r, \theta, z),\ y(r, \theta, z), z(r, \theta, z))$$

Composite Function Formulation

Perhaps a more intuitive way to examine the Chain Rule is to consider the function as a composition of multiple functions.

$$h(U) = f(g(U))$$

where $\quad U = \begin{pmatrix} r \\ \theta \\ z \end{pmatrix}$

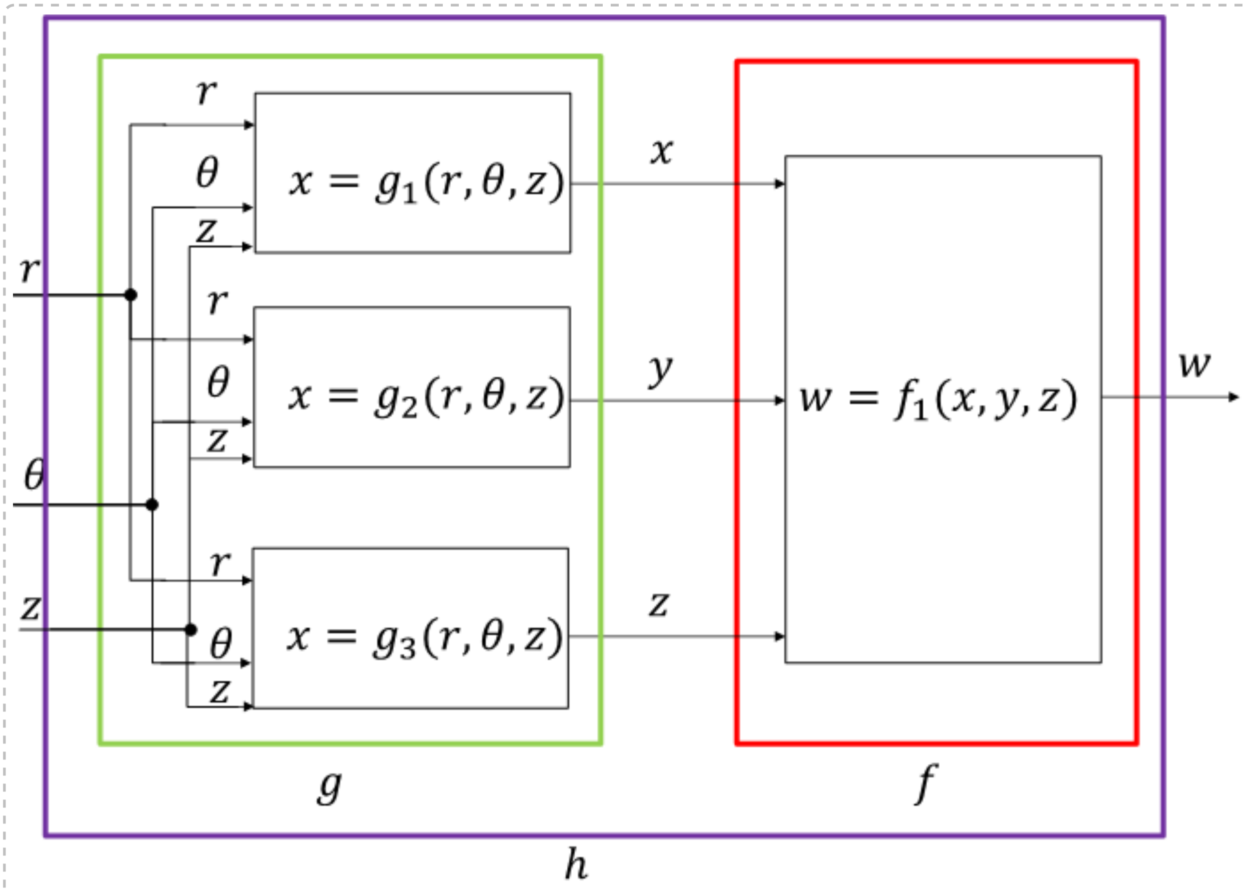
Sometimes written as

$$h = f \circ g$$

Using the example above

$$w = f(g(r,\ \theta,\ z))$$

This is easiest to visualize using a block diagram as shown below



The goals is to compute

$$\begin{pmatrix} \frac{\partial w}{\partial r} \\ \frac{\partial w}{\partial \theta} \\ \frac{\partial w}{\partial z} \end{pmatrix} = ??$$

In this case we have

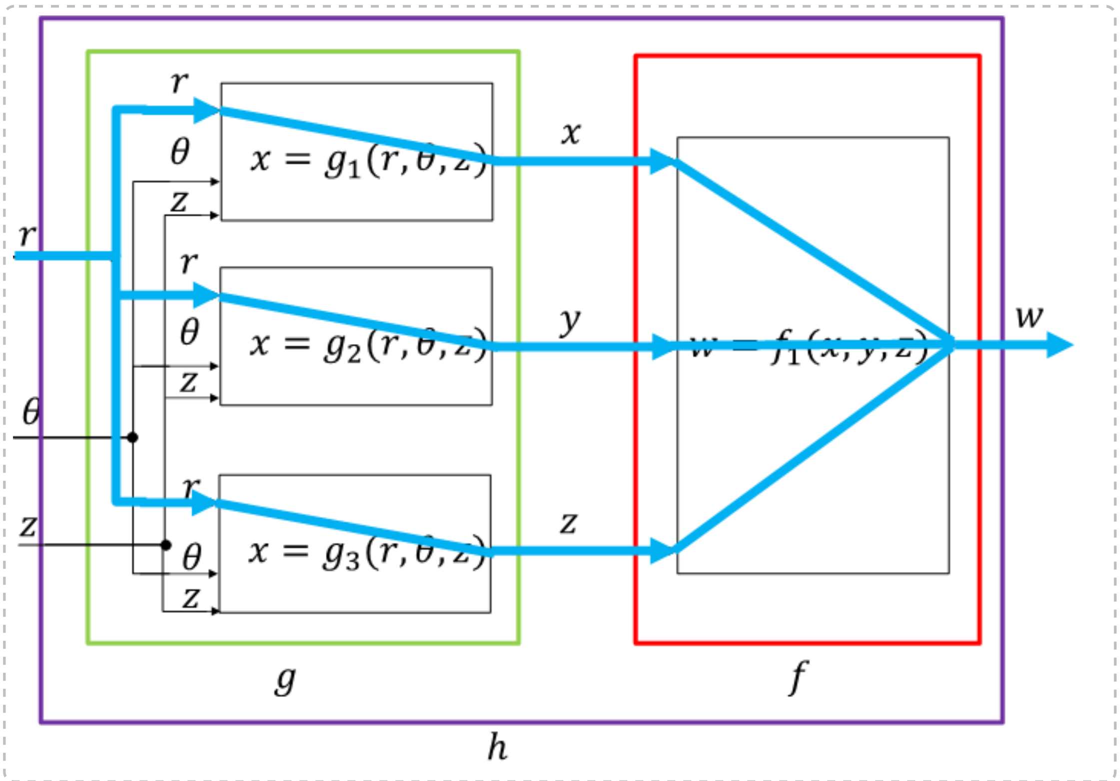
$$g(U) = g(r,\ \theta,\ z) = \begin{pmatrix} g_1(r,\ \theta,\ z) \\ g_2(r,\ \theta,\ z) \\ g_3(r,\ \theta,\ z) \end{pmatrix} = \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \\ z \end{pmatrix}$$

$$f(X) = f(x,\ y,\ z) = f_1(x,\ y,\ z) = 2\,x^2\,y - 4\,y^2 + z^3$$

```
In[ ]:= f[x_, y_, z_] = {{2 x^2 y - 4 y^2 + z^3}};

g[r_, θ_, z_] = (r Cos[θ]
                 r Sin[θ]
                 z);
```

We can compute the sensitivity of *w* w.r.t. *r* by examining the pathways through the block diagram that lead to *w*



We see that changing/varying *r* will vary *x*, *y*, and *z*. The sensitivity of *x*, *y*, and *z* to variations in *r* are given by

$$\frac{\partial x}{\partial r} = \frac{d}{dr}[r\cos(\theta)] = \cos(\theta)$$

$$\frac{\partial y}{\partial r} = \frac{\partial}{\partial r}[r\sin(\theta)] = \sin(\theta)$$

$$\frac{\partial z}{\partial r} = \frac{\partial}{\partial r}[z] = 0$$

```
In[]:= temp = D[g[r, θ, z], r];
       dxdr = temp[[1, 1]]
       dydr = temp[[2, 1]]
       dzdr = temp[[3, 1]]
Out[]= Cos[θ]
Out[]= Sin[θ]
Out[]= 0
```

We see that f_1 is a function of x, y, and z which we have just established will change with variations in r. As such, the overall effect on the output variable w w.r.t. variations in r is then given by

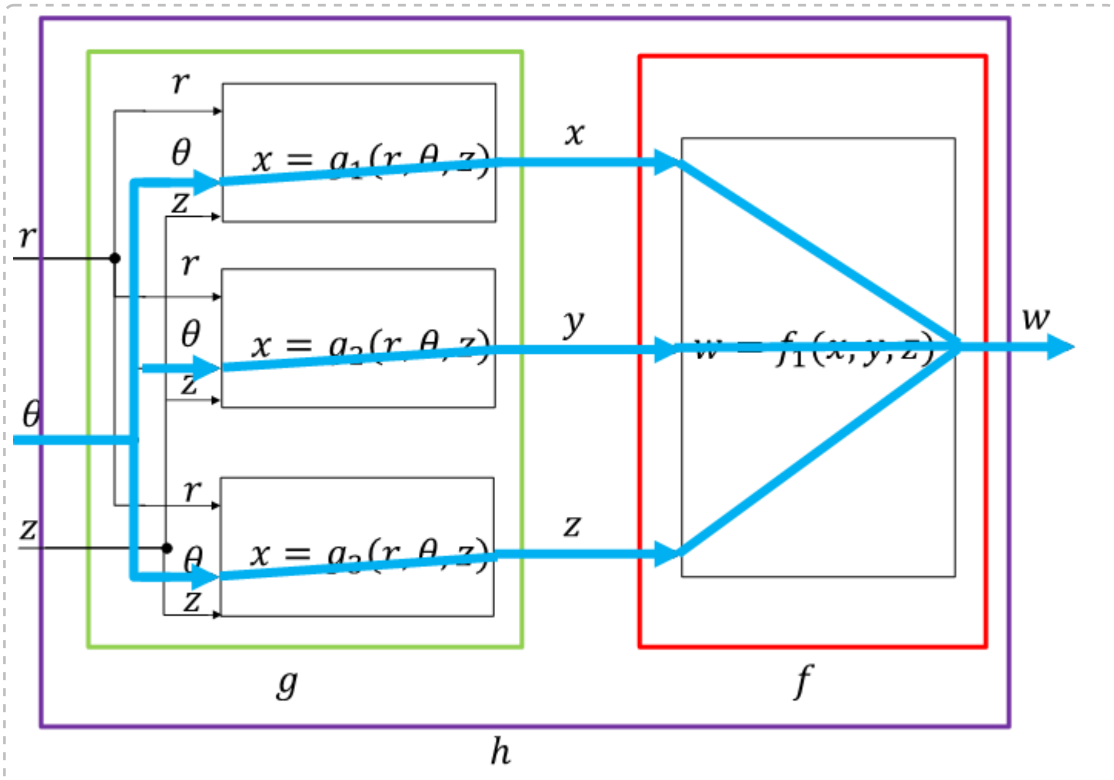
$$\frac{\partial w}{\partial r} = \frac{\partial x}{\partial r}\frac{\partial w}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial w}{\partial y} + \frac{\partial z}{\partial r}\frac{\partial w}{\partial z} \qquad \textbf{(Eq.2.a)}$$

We previously computed $\frac{\partial w}{\partial x}$, $\frac{\partial w}{\partial y}$, and $\frac{\partial w}{\partial z}$ so we have

$$\frac{\partial w}{\partial r} = (\cos(\theta))*(4\,x\,y) + (\sin(\theta))*\left(2\,x^2 - 8\,y\right) + (0)*\left(3\,z^2\right) \qquad \text{recall: } x = r\cos(\theta),\ y = r\sin(\theta),\ z = z$$

```
In[]:= dwdr = dxdr * dwdx + dydr * dwdy + dzdr * dwdz /.
          {x → x[r, θ, z], y → y[r, θ, z], z → z[r, θ, z]} // Simplify
Out[]= 2 r (3 r Cos[θ]^2 - 4 Sin[θ]) Sin[θ]
```

In a similar fashion, we can examine how changes in θ affects the output w. Examining the pathways through the block diagram that lead to w yields the following



We see that changing/varying θ will vary x, y, and z. The sensitivity of x, y, and z to variations in θ are given by

$$\frac{\partial x}{\partial \theta} = \frac{d}{d\theta}[r\cos(\theta)] = -r\sin(\theta)$$

$$\frac{\partial y}{\partial \theta} = \frac{\partial}{\partial \theta}[r\sin(\theta)] = r\cos(\theta)$$

$$\frac{\partial z}{\partial \theta} = \frac{\partial}{\partial \theta}[z] = 0$$

In[ ]:=
```
temp = D[g[r, θ, z], θ];
dxdθ = temp[[1, 1]]
dydθ = temp[[2, 1]]
dzdθ = temp[[3, 1]]
```

Out[ ]= -r Sin[θ]

Out[ ]= r Cos[θ]

Out[ ]= 0

We see that f_1 is a function of x, y, and z which we have just established will change with variations in θ. As such, the overall effect on the output variable w w.r.t. variations in θ is then given by

$$\frac{\partial w}{\partial \theta} = \frac{\partial x}{\partial \theta}\frac{\partial w}{\partial x} + \frac{\partial y}{\partial \theta}\frac{\partial w}{\partial y} + \frac{\partial z}{\partial \theta}\frac{\partial w}{\partial z} \qquad \textbf{\textit{(Eq.2.b)}}$$

We previously computed $\frac{\partial w}{\partial x}$, $\frac{\partial w}{\partial y}$, and $\frac{\partial w}{\partial z}$ so we have

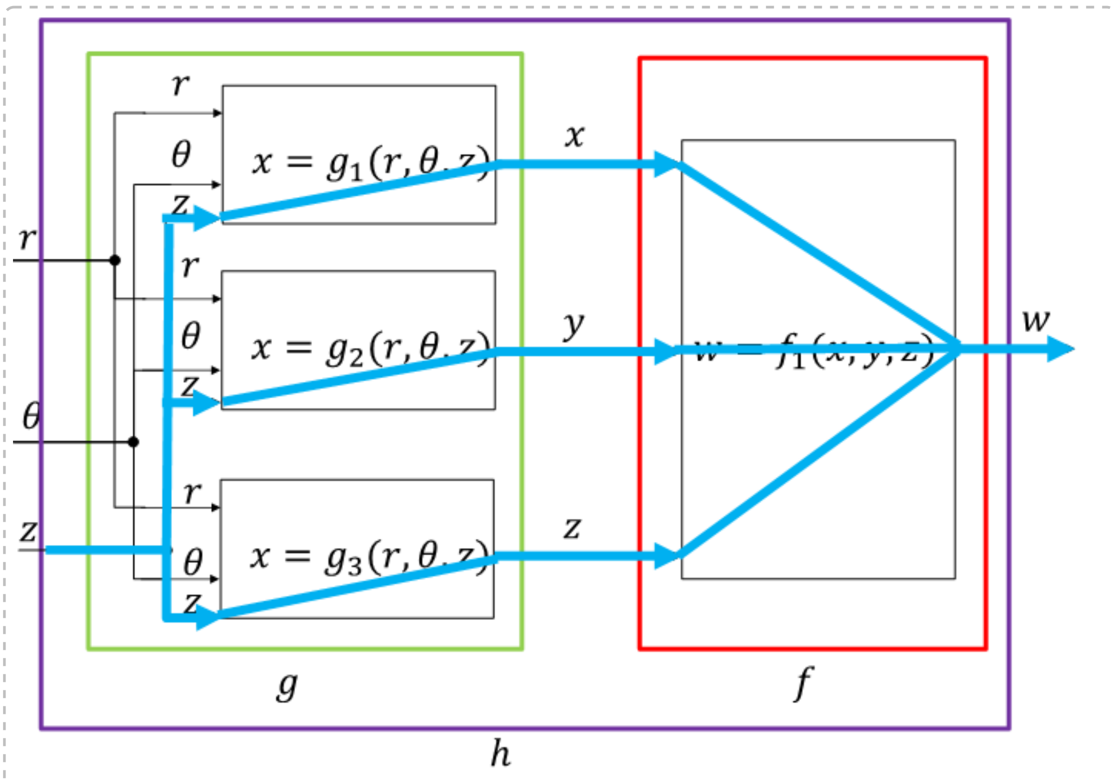
$$\frac{\partial w}{\partial \theta} = (-r\sin(\theta)) * (4\,x\,y) + (r\cos(\theta)) * \left(2\,x^2 - 8\,y\right) + (0) * \left(3\,z^2\right)$$ recall:

$x = r\cos(\theta)$, $y = r\sin(\theta)$, $z = z$

```
In[∘]:= dwdθ = dxdθ * dwdx + dydθ * dwdy + dzdθ * dwdz /.
            {x → x[r, θ, z], y → y[r, θ, z], z → z[r, θ, z]} // Simplify
```

```
Out[∘]= r^2 Cos[θ] (-r + 3 r Cos[2 θ] - 8 Sin[θ])
```

In a similar fashion, we can examine how changes in z affects the output w. Examining the pathways through the block diagram that lead to w yields the following



We see that changing/varying z will vary x, y, and z. The sensitivity of x, y, and z to variations in z are given by

$$\frac{\partial x}{\partial z} = \frac{d}{d\,z}[r\cos(\theta)] = 0$$

$$\frac{\partial y}{\partial z} = \frac{\partial}{\partial z}[r\sin(\theta)] = 0$$

$$\frac{\partial z}{\partial z} = \frac{\partial}{\partial z}[z] = 1$$

In[∘]:=
```
temp = D[g[r, θ, z], z];
dxdz = temp[[1, 1]]
dydz = temp[[2, 1]]
dzdz = temp[[3, 1]]
```

Out[∘]= 0

Out[∘]= 0

Out[∘]= 1

We see that f_1 is a function of x, y, and z which we have just established will change with variations in z. As such, the overall effect on the output variable w w.r.t. variations in z is then given by

$$\frac{\partial w}{\partial z} = \frac{\partial x}{\partial z}\frac{\partial w}{\partial x} + \frac{\partial y}{\partial z}\frac{\partial w}{\partial y} + \frac{\partial z}{\partial z}\frac{\partial w}{\partial z} \qquad \textbf{\textit{(Eq.2.c)}}$$

We previously computed $\frac{\partial w}{\partial x}$, $\frac{\partial w}{\partial y}$, and $\frac{\partial w}{\partial z}$ so we have

$$\frac{\partial w}{\partial z} = (0)*(4\,x\,y) + (0)*\left(2\,x^2 - 8\,y\right) + (1)*\left(3\,z^2\right) \qquad \text{recall: } x = r\cos(\theta),\ y = r\sin(\theta),\ z = z$$

In[∘]:=
```
dwdz = dxdz * dwdx + dydz * dwdy + dzdz * dwdz /.
   {x → x[r, θ, z], y → y[r, θ, z], z → z[r, θ, z]} // Simplify
```

Out[∘]= $3\,z^2$

So we have

$$\begin{pmatrix} \frac{\partial w}{\partial r} \\ \frac{\partial w}{\partial \theta} \\ \frac{\partial w}{\partial z} \end{pmatrix} = \begin{pmatrix} 2\,r\left(3\,r\,\mathrm{Cos}[\theta]^2 - 4\,\mathrm{Sin}[\theta]\right)\mathrm{Sin}[\theta] \\ r^2\,\mathrm{Cos}[\theta]\,(-r + 3\,r\,\mathrm{Cos}[2\,\theta] - 8\,\mathrm{Sin}[\theta]) \\ 3\,z^2 \end{pmatrix}$$

Note that we can again obtain this by simply using brute force techniques

In[∘]:=
```
(*Brute force composite function*)
Clear[h]
temp = f[x[r, θ, z], y[r, θ, z], z[r, θ, z]];
h[r_, θ_, z_] = temp[[1, 1]]

(*Compute derivatives*)
Print["Derivatives"]
dhdr = D[h[r, θ, z], r]
dhdθ = D[h[r, θ, z], θ]
dhdz = D[h[r, θ, z], z]

(*Compare with chain rule*)
Print["Compare to Chain Rule"]
dwdr == dhdr // Simplify
dwdθ == dhdθ // Simplify
dwdz == dhdz // Simplify
```

Out[∘]= $z^3 + 2\,r^3\,\mathrm{Cos}[\theta]^2\,\mathrm{Sin}[\theta] - 4\,r^2\,\mathrm{Sin}[\theta]^2$

```
Derivatives
```

Out[∘]= $6\,r^2\,\mathrm{Cos}[\theta]^2\,\mathrm{Sin}[\theta] - 8\,r\,\mathrm{Sin}[\theta]^2$

Out[∘]= $2\,r^3\,\mathrm{Cos}[\theta]^3 - 8\,r^2\,\mathrm{Cos}[\theta]\,\mathrm{Sin}[\theta] - 4\,r^3\,\mathrm{Cos}[\theta]\,\mathrm{Sin}[\theta]^2$

Out[∘]= $3\,z^2$

```
Compare to Chain Rule
```

Out[∘]= True

Out[∘]= True

Out[∘]= True

Jacobian Matrix Formulation

Let us revisit Eq.2.a, Eq.2.b, and Eq.2.b, repeated here for convenience

$$\frac{\partial w}{\partial r} = \frac{\partial x}{\partial r}\frac{\partial w}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial w}{\partial y} + \frac{\partial z}{\partial r}\frac{\partial w}{\partial z} \qquad \textit{(Eq.2.a)}$$

$$\frac{\partial w}{\partial \theta} = \frac{\partial x}{\partial \theta}\frac{\partial w}{\partial x} + \frac{\partial y}{\partial \theta}\frac{\partial w}{\partial y} + \frac{\partial z}{\partial \theta}\frac{\partial w}{\partial z} \qquad \textit{(Eq.2.b)}$$

$$\frac{\partial w}{\partial z} = \frac{\partial x}{\partial z}\frac{\partial w}{\partial x} + \frac{\partial y}{\partial z}\frac{\partial w}{\partial y} + \frac{\partial z}{\partial z}\frac{\partial w}{\partial z} \qquad \textit{(Eq.2.c)}$$

Note that we can write this in matrix form as

$$\begin{pmatrix} \frac{\partial w}{\partial r} \\ \frac{\partial w}{\partial \theta} \\ \frac{\partial w}{\partial z} \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial r}\frac{\partial w}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial w}{\partial y} + \frac{\partial z}{\partial r}\frac{\partial w}{\partial z} \\ \frac{\partial x}{\partial \theta}\frac{\partial w}{\partial x} + \frac{\partial y}{\partial \theta}\frac{\partial w}{\partial y} + \frac{\partial z}{\partial \theta}\frac{\partial w}{\partial z} \\ \frac{\partial x}{\partial z}\frac{\partial w}{\partial x} + \frac{\partial y}{\partial z}\frac{\partial w}{\partial y} + \frac{\partial z}{\partial z}\frac{\partial w}{\partial z} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} & \frac{\partial z}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} & \frac{\partial z}{\partial \theta} \\ \frac{\partial x}{\partial z} & \frac{\partial y}{\partial z} & \frac{\partial z}{\partial z} \end{pmatrix} \begin{pmatrix} \frac{\partial w}{\partial x} \\ \frac{\partial w}{\partial y} \\ \frac{\partial w}{\partial z} \end{pmatrix}$$

$$\begin{pmatrix} \frac{\partial w}{\partial r} \\ \frac{\partial w}{\partial \theta} \\ \frac{\partial w}{\partial z} \end{pmatrix}^T = \begin{pmatrix} \frac{\partial w}{\partial x} \\ \frac{\partial w}{\partial y} \\ \frac{\partial w}{\partial z} \end{pmatrix}^T \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} & \frac{\partial z}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} & \frac{\partial z}{\partial \theta} \\ \frac{\partial x}{\partial z} & \frac{\partial y}{\partial z} & \frac{\partial z}{\partial z} \end{pmatrix}^T$$

$$= \begin{pmatrix} \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial z} \end{pmatrix}$$

As we see, these aforementioned calculations of derivatives of the composite function involve the partial derivatives of the individual functions g and f. Recall from our discussion of the Jacobian Matrix (see YouTube video entitled 'The Jacobian Matrix' at https://youtu.be/QexBVGVM690) that we can write

$$\begin{pmatrix} \frac{\partial w}{\partial r} \\ \frac{\partial w}{\partial \theta} \\ \frac{\partial w}{\partial z} \end{pmatrix}^T = J_f \, J_g$$

where $J_f = \begin{pmatrix} \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{pmatrix}$ (Jacobian of function f)

$$J_g = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial z} \end{pmatrix}$$ (Jacobian of function g)

```
In[]:= Jf = ( dwdx  dwdy  dwdz );

       Jg = ( dxdr  dxdθ  dxdz
              dydr  dydθ  dydz
              dzdr  dzdθ  dzdz );

       dwdU = Jf.Jg /. {x → x[r, θ, z], y → y[r, θ, z], z → z[r, θ, z]} // Simplify;
       dwdU // MatrixForm

       Print["Check against chain rule"]
       dwdU[[1, 1]] == dwdr
       dwdU[[1, 2]] == dwdθ
       dwdU[[1, 3]] == dwdz
```

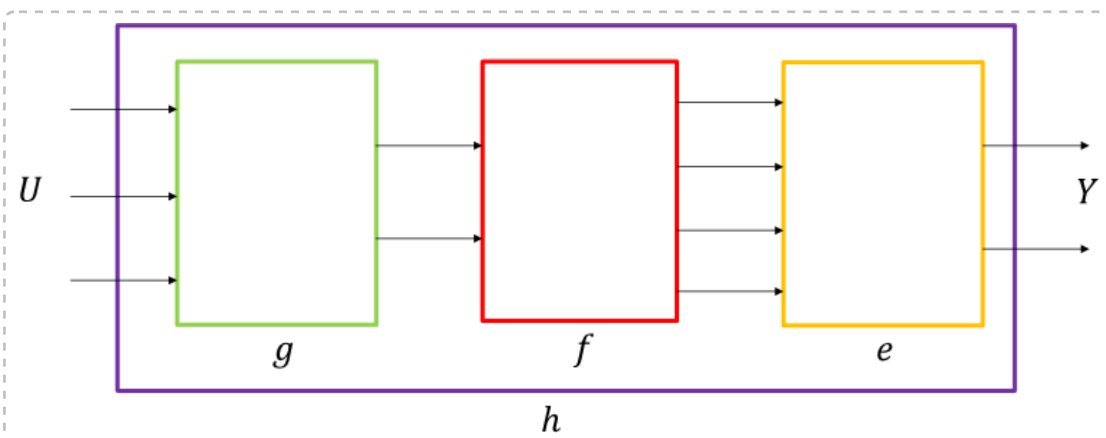
Out[]//MatrixForm=

$$\left(\begin{array}{ccc} 2\, r \left(3\, r \cos[\theta]^2 - 4 \sin[\theta]\right) \sin[\theta] & r^2 \cos[\theta] \left(-r + 3\, r \cos[2\,\theta] - 8 \sin[\theta]\right) & 3\, z^2 \end{array}\right)$$

```
Check against chain rule
Out[]= True
Out[]= True
Out[]= True
```

Chain Rule (Multiple Input Multiple Output)

We can extend this concept to a system with multiple input and multiple outputs and multiple layers of composite functions. For example, consider the following composite function



This has multiple inputs U and multiple outputs Y. The Chain Rule can easily be expressed using the Jacobian matrix formulation to capture all the sensitivities (AKA partial derivatives) as

$$\frac{\partial Y}{\partial U} = J_e\, J_f\, J_g$$

This has applications in areas such as Neural Networks (https://youtu.be/i2fmaabIs5w) and the Backpropagation Algorithm.