# Derivatives, partial derivatives, influence diagrams, and other magical beasts

## 1. Defining the derivative

Let $y = f(x)$ be a relation between two variables, $y$ and $x$. If $f(x)$ is continuous and differentiable, any small perturbation of $x$ will result in a small perturbation of $y$. We define the *derivative* of $y$ with respect to $y$ as the *multiplier* alpha that relates a miniscule perturbation $\delta x$ of $x$ to the resulting perturbation $\delta y$ of $y$.

$$\delta y = \alpha . \delta x \tag{1}$$

Figure 1 illustrates this relation.

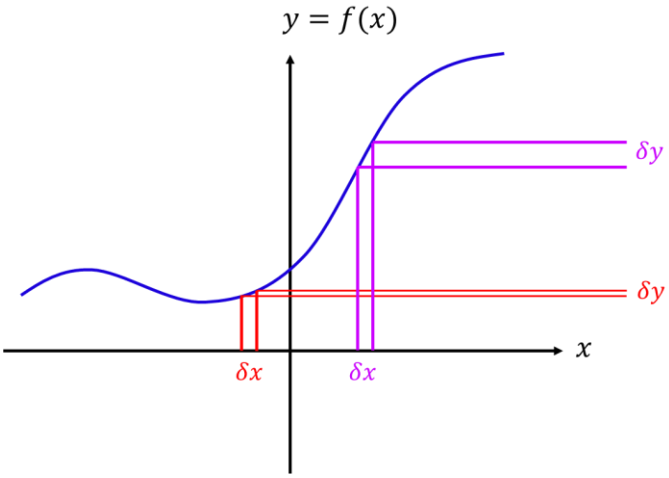


Figure 1. A small increment $\delta x$ will result in a corresponding perturbation $\delta y$ of $y$. The value of $\delta y$ will depend on the location of $x$ of the perturbation $\delta x$. In the figure, the $\delta y$ (in red) resulting from the perturbation $\delta x$ at the red location is entirely different from the perturbation n $\delta y$ at the purple location.

As evident from the figure, the change $\delta y$ resulting from a perturbation $\delta x$ generally depends on the specific location around which $x$ is perturbed, so in fact, the value alpha is better written as $\alpha(x)$, to indicate that it is in fact dependent on $x$.

When $y$ and $x$ are both scalars, we will often represent $\alpha(x)$ as $f'(x)$, a representation that indicates both the relation to the original function $f(.)$ and the dependence on $x$ itself, and rewrite the perturbation relation of Equation 1 either as

$$f'(x) = \frac{dy}{dx} \tag{2}$$

## 1.1. Derivatives for scalar functions of vectors

When $x$ and/or $y$ are vectors (e.g for the output of vector activation, e.g. a softmax, where input and output are both vectors, or for a divergence function, which is a scalar value that is a function of the vector of output variables in the network), the above notation is not usable. Let us first consider the case when $\mathbf{x}$ is a (column) vector and $y$ is a scalar. The acceptable notation for the derivative Equation remains of the form

$$\delta y = \nabla_{\mathbf{x}} f(\mathbf{x}).\delta\mathbf{x} \tag{3}$$

Here, we use the bolded symbol $\mathbf{x}$ to indicate that it is a vector. $\nabla_{\mathbf{x}} f(\mathbf{x})$ is the derivative of $f(\mathbf{x})$ with respect to $\mathbf{x}$, computed at the location $\mathbf{x}$. We may also use the short-hand notation of $\nabla_{\mathbf{x}} y$ to represent it (as we will in Section 1.2). Note that this is identical to Equation 1, which is the original definition of a derivative.

As explained in class, for the dimensions of the above equation to be valid, the product on the right hand side must return a scalar (since $y$ and therefore $\delta y$ are scalars). Since, by our convention, $\mathbf{x}$ is a column vector, $\delta\mathbf{x}$ too is a column vector. Consequently, it is obvious that $\nabla_{\mathbf{x}} f(\mathbf{x})$ must be a row vector of the same size as the *transpose* of $\mathbf{x}$. Specifically, if $\mathbf{x}$ is a $D \times 1$ vector, $\nabla_{\mathbf{x}} f(\mathbf{x})$ will have the form

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_D} \end{bmatrix} \tag{4}$$

where $\frac{\partial y}{\partial x_i}$ is a *partial* derivative of $f(\mathbf{x})$ which quantifies how much $y$ changes when the specific component $x_i$ is perturbed, while keeping all other components of $\mathbf{x}$ fixed, i.e.

$$\delta y = \frac{\partial y}{\partial x_i}\delta x_i, \quad \text{given } \delta x_j = 0 \; \forall \; j \neq i \tag{5}$$

(the symbol "∀" stands for "for all". We could write it in English, but it doesn't look as profound that way ☺).

The term "partial" in the name indicates the fact that we are only *partially* considering the influence of $x_i$ on $y$. This will not be immediately obvious from Equations 4 or 5, but should become clearer by the time you're done reading this document.

## 1.2. Derivatives for vector functions of vectors

When both $\mathbf{x}$ and $\mathbf{y}$ are (column) vectors the derivative equation becomes

$$\delta\mathbf{y} = \nabla_{\mathbf{x}} f(\mathbf{x}).\delta\mathbf{x} \tag{6}$$

If $\mathbf{y}$ is a $K \times 1$ vector, and $\mathbf{x}$ is a $D \times 1$ vector, it is clear from Equation 6 that $\nabla_{\mathbf{x}} f(\mathbf{x})$ must be a $K \times D$ matrix. The derivative matrix $\nabla_{\mathbf{x}} f(\mathbf{x})$ is often referred to as the *Jacobian* of $f(.)$ and may be designated instead as $\mathrm{J}_{\mathbf{x}} f(\mathbf{x})$.

Representing the individual components of the vectors $\mathbf{x}$ and $\mathbf{y}$ as $x_i$ and $y_j$ respecitvely, the Jacobian is given by

$$\mathrm{J}_{\mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_D} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_D} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_K}{\partial x_1} & \frac{\partial y_K}{\partial x_2} & \cdots & \frac{\partial y_K}{\partial x_D} \end{bmatrix} \tag{7}$$

where $\frac{\partial y_i}{\partial x_j}$ is the partial derivative of as $y_i$ with respect to $x_j$.

Note that there is nothing magical about the Jacobian. The $i^{\text{th}}$ row of the matrix is simply the derivative of $y_i$ with respect to $\mathbf{x}$ . In fact, the derivative Equation 6 simply treats the column vector $\mathbf{y}$ as a vertical stacking of $K$ variables (namely $y_1 \ldots y_K$), and the derivative of Equation 7 too is simply the vertical stacking of the derivatives of each of these $K$variables with respect to $\mathbf{x}$.

### *1.3. Derivatives for functions of matrices and tensors*

When $\mathbf{x}$ is a higher order algebraic term, like a matrix or tensor, its entries can be "unwound" to reshape it as a vector. So also, if $\mathbf{y}$ is a higher-order term like a matrix or tensor, its entries too can be reshaped into a vector. Consequently, the rules of section 1.2 apply.

Often, we will perform additional simplifications to produce simpler formulae, like the derivative of the divergence w.r.t. weights matrices for networks. We will not address that topic in this document.

*A key take-away from this section is that if you know how to compute derivatives for a scalar function of vector variables, you can generalize that knowledge to compute derivatives for vector and higher-order functions of vector, matrix and other higher-order input variables.*

## 2. Partial and full derivatives, and influence diagrams

We now arrive at the main point of this note.

## 2.1. The chain rule

Consider the case when $y$ is a function of several variables $x_1, x_2, \ldots, x_D$. We can write

$$y = f(x_1, x_2, \ldots, x_D) \tag{8}$$

Note that this equation is identical to the case of section 1.2, except that instead of writing the variables $x_1, x_2, \ldots, x_D$ as a vector, we are writing them out separately. We will assume that all variables are scalars in this section. We leave the generalization to vectors to the reader, with the assurance that it is, in fact, quite straight forward.

The derivative equation for $y$ can now be written as

$$\delta y = \frac{\partial y}{\partial x_1}\delta x_1 + \frac{\partial y}{\partial x_2}\delta x_2 + \cdots + \frac{\partial y}{\partial x_D}\delta x_D \tag{9}$$

The term $\frac{\partial y}{\partial x_i}$ is known as a *partial* derivative. It specifies how much $y$ changes in response to an infinitesimal perturbation of $x_i$ *when the remaining* $x_{j\neq i}$ *are not perturbed.* To reiterate,

$$\delta y = \frac{\partial y}{\partial x_i}\delta x_i, \quad \text{given } \delta x_j = 0 \;\forall\; j \neq i \tag{10}$$

Why do we call this a *partial* derivative? Because we're assuming that perturbing $x_i$ has no influence on the remaining variables. This may not necessarily be true. In the examples below, we will use $x_1$ as an illustrative example of $x_i$, but the discussion clearly generalizes. Consider the case where $x_2 = g(x_1)$, i.e. is a function of $x_1$. A small perturbation of $x_1$ will also naturally result in a small perturbation of $x_2$, which in turn will have an influence on $y$. The computation of Equation 10 ignores this influence.

What then is the *full* derivative of $y$ w.r.t. (say) $x_1$? The full derivative computes the *total* influence of $x_1$ on $y$ and is written as $\frac{dy}{dx_1}$. The derivative relation, once again, is given by

$$\delta y = \frac{dy}{dx_1}\delta x_1 \tag{11}$$

and includes both *direct* and *indirect* influences of $x_1$ on $y$ (as opposed to the partial derivative, which only includes direct influences). To compute the full derivative, you must consider the influence of $x_1$ on the other variables. Thus, if $x_2$ is a function of $x_1$, then the small perturbation $\delta x_2$ in Equation 9 is also actually a function of $\delta x_2$, and can be written as

$$\delta x_2 = \frac{dx_2}{dx_1}\delta x_1, \tag{12}$$

where $\frac{dx_2}{dx_1}$ is the *full* derivative of $x_2$ w.r.t $x_1$. Thus the actual derivative equation for y, obtained by plugging Equation 12 into Equation 9 is

$$\delta y = \frac{\partial y}{\partial x_1}\delta x_1 + \frac{\partial y}{\partial x_2}\frac{dx_2}{dx_1}\delta x_1 + \frac{\partial y}{\partial x_3}\delta x_3 + \cdots + \frac{\partial y}{\partial x_D}\delta x_D \tag{13}$$

If *all* $x_i$ are dependent on $\delta x_1$we have the following situation : $x_i = g_i(x_1)$ then, following the same logic, we can write

$$\delta y = \frac{\partial y}{\partial x_1}\delta x_1 + \frac{\partial y}{\partial x_2}\frac{dx_2}{dx_1}\delta x_1 + \frac{\partial y}{\partial x_3}\frac{dx_3}{dx_1}\delta x_1 + \cdots + \frac{\partial y}{\partial x_D}\frac{dx_D}{dx_D}\delta x_1 \tag{14}$$

or, grouping terms together,

$$\delta y = \left(\frac{\partial y}{\partial x_1} + \frac{\partial y}{\partial x_2}\frac{dx_2}{dx_1} + \cdots + \frac{\partial y}{\partial x_D}\frac{dx_D}{dx_1}\right)\delta x_1 \tag{15}$$

Comparing Equation 15 to Equation 11, we determine that the *full* derivative of $y$ w.r.t. $x_1$ is given by

$$\frac{dy}{dx_1} = \frac{\partial y}{\partial x_1} + \frac{\partial y}{\partial x_2}\frac{dx_2}{dx_1} + \cdots + \frac{\partial y}{\partial x_D}\frac{dx_D}{dx_1} \tag{16}$$

(once again, note which terms are partial derivatives and which are full derivatives in Equation 16. The full derivatives w.r.t. $x_i$are required to derive $\delta x_i$, the *total* variation of $x_i$ in response to a perturbation $\delta x_1$ of $x_1$).

When none of the $x_{j\neq 1}$ are themselves functions of $x_1$ then $\frac{dx_j}{dx_1} = 0$ for $j \neq 1$ and the partial derivative of $y$ w.r.t. $x_1$ is exactly equal to the *full* derivative, i.e. $\frac{dy}{dx_1} = \frac{\partial y}{\partial x_1}$.

You will often find a shorthand notation of the following kind when $y$ is a function of multiple variables, which are themselves interdependent. For instance, as in the above example, if $y$ is a function of $x_1, x_2, \ldots, x_D$, where $x_i$ in turn is a function of $x_1$, then you may find the relation written directly as

$$y = f\left(x_1, g_2(x_1), g_3(x_1), \ldots, g_D(x_1)\right) \tag{17}$$

and the derivative relation written as

$$\delta y = \frac{\partial y}{\partial x_1}\delta x_1 + \frac{\partial y}{\partial g_2}\frac{dg_2}{dx_1}\delta x_1 + \cdots + \frac{\partial y}{\partial g_D}\frac{dg_D}{dx_1}\delta x_1 \tag{18}$$

$$\frac{dy}{dx_1} = \frac{\partial y}{\partial x_1} + \frac{\partial y}{\partial g_2}\frac{dg_2}{dx_1} + \cdots + \frac{\partial y}{\partial g_D}\frac{dg_D}{dx_1} \tag{19}$$

Equations 17-19 are in fact no different from simply writing $y = f(x_1, x_2, \ldots, x_D)$ and $x_i = g_i(x_1)$, and using Equations 14 or 16. When in doubt, we recommend rewriting equations using the latter approach since it usually gives us a clearer view of the relationships and derivative relations.

A special case of Equation 17 is when all $x_1, x_2, \ldots, x_D$ are functions of a common variable $x_i = x$ i.e. $x_i = g_i(x)$, for all $x_i$.

$$y = f\big(g_1(x), g_2(x), g_3(x), \ldots, g_D(x)\big) \tag{20}$$

We recognize immediately that Equation 17 is merely a special case of Equation 20, where $g_1(x)$ is the identity function. The derivative relations can be written as

$$\delta y = \frac{\partial y}{\partial g_1}\frac{dg_1}{dx}\delta x + \frac{\partial y}{\partial g_2}\frac{dg_2}{dx}\delta x + \cdots + \frac{\partial y}{\partial g_D}\frac{dg_D}{dx}\delta x \tag{21}$$

$$\frac{dy}{dx} = \frac{\partial y}{\partial g_1}\frac{dg_1}{dx} + \frac{\partial y}{\partial g_2}\frac{dg_2}{dx} + \cdots + \frac{\partial y}{\partial g_D}\frac{dg_D}{dx} \tag{22}$$

Once again, it's obvious from inspection of Equations 21 and 22 that they are almost identical to Equations 18 and 19; all we have done is to replace the variable $x_i$ by the function $g_i$ that computes it. Here too, please note which terms are partial derivatives, and which ones are full derivatives. We will understand the rationale behind this in the next section hopefully.

## *2.2. Influence Diagrams*

The equations in the previous section may be better understood if we *visualize* the dependencies between variables. We will attempt to do so below.

### *2.2.1. The simple chain rule as an influence diagram*

As a warm up, let's consider the well-known case $y = f(x_1, x_2, \ldots, x_D)$.

We can draw a diagram of influences of the various variables in this equation. This is shown in Figure 2 below. The arrows show that each of $x_1, x_2, \ldots, x_D$ have an influence on $y$.

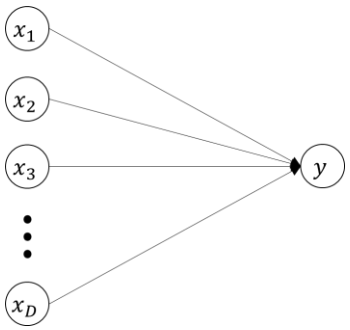


Figure 2. Influence diagram for $y = f(x_1, x_2, \ldots, x_D)$

From the definition of the derivative in Equation 9 (or equivalently, the definition as given by Equations 3 and 4), the derivative rule for the relationship pictured by the above influence diagram is

$$\delta y = \frac{\partial y}{\partial x_1}\delta x_1 + \frac{\partial y}{\partial x_2}\delta x_2 + \cdots + \frac{\partial y}{\partial x_D}\delta x_D \tag{23}$$

This reiteration of Equation 9 shows exactly how perturbations of the inputs influence the output. Each of the input variables may be perturbed (by $\delta x_i$). This perturbation results in a variable-specific change of the output, equal to $\frac{\partial y}{\partial x_i}\delta x_i$. The overall perturbation in $y$ is the *cumulative sum* of the output perturbations due to each of the input variables.

The derivative rule uses the *partial derivative* $\frac{\partial y}{\partial x_i}$ to quantify the influence of variable $x_i$upon the output since we aren't actually considering whether any of $x_1, x_2, \ldots, x_D$ depend on each other in turn.

Equation 23 remains valid *regardless of any dependencies between* $x_1, x_2, \ldots, x_D$. How any dependencies will influence the overall perturbation $\delta y$ is *entirely a function of how the increments* $\delta x_i$ *relate to one another*.

It is worth noting here that if $x_1, x_2, \ldots, x_D$ are fully independent of one another and have no dependencies among themselves, then we could replace the partial derivatives in Equation 23 by the *full* derivatives $\frac{dy}{dx_i}$ in Equation 23, and it would remain correct.

So what are these partial and full derivatives, and how do we use them? We will explain this below.

### *2.2.2. Partial and full derivatives through influence diagrams*

Remember that the derivative of a variable $y$ with respect to a variable $x$ simply quantifies how much $y$ varies in response to a minute change of $x$. $x$ may influence $y$ either directly, or indirectly, through its influence on other variables which in turn influence $y$. The figures below show the two situations.


Figure 2a: $x$ only has a direct influence on $y$

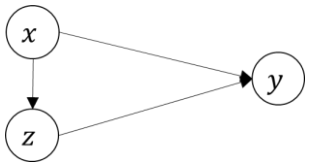


Figure 2b: $x$ has both direct and indirect influence on $y$

The figures illustrate *influence*. In the Figure 2a, $x$ influences $y$ directly as shown. The underlying relation can be written in the form $y = f(x)$. Since there is only a direct link between $x$ and $y$, the derivative relation between the two is given by

$$\delta y = \frac{dy}{dx}\delta x \tag{24}$$

where $\frac{dy}{dx}$ is the *full* derivative of $y$ w.r.t. $x$.

In Figure 2b, $x$ influences $y$ along multiple paths. It has a direct influence on $y$. It also influences $z$. $z$, in turn, also influences $y$. This influence diagram represents a functional relation of the form $z = g(x)$ , $y = f(x, z)$, which could be compacted into $y = f(x, g(x))$.

Since $x$ has both direct and indirect influences on $y$ we can define both partial and full derivatives of $y$ w.r.t. $x$.

Perturbing $x$ in the influence diagram of Figure 2b influences $y$ along two paths as shown in Figure 3a below. The two influences accumulate additively. The *full* derivative of $y$ w.r.t. $x$ must consider both paths.

The *partial* derivative $\frac{\partial y}{\partial x}$ describes the derivative of $y$ w.r.t $x$ if only the value over the direct link from $x$ to $y$ were perturbed as shown in Fig 3b below. In essence, when computing this partial derivative, we "snip" the dependence of other variables that influence $y$. Note that this is a *partial* derivative, since there are other indirect paths from $x$ to $y$ that are being ignored (snipped) in computing this derivative. The resulting derivative of $y$ w.r.t. $x$ is the *partial* derivative of $y$ w.r.t. $x$. The derivative relation is given by

$$\delta y = \frac{dy}{dx}\delta x \tag{25}$$

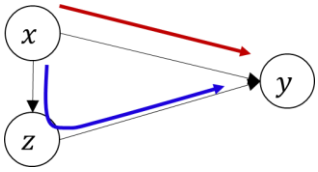


Figure 3a: The two paths through which $x$ influences $y$. The red line the path of direct influence. The blue line shows the indirect influence, through $z$.

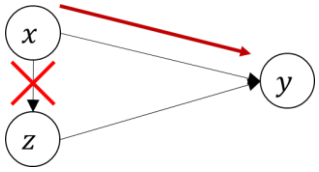


Figure 3b: The influences considered when computing the partial derivative of $y$ w.r.t $x$.

The *full* derivative relation of $y$ to $x$ is in fact given by

$$\delta y = \left(\frac{\partial y}{\partial x} + \frac{\partial y}{\partial z}\frac{dz}{dx}\right)\delta x \tag{26}$$

and includes the partial derivative as a component. Thus the full derivative of $y$ with respect to $x$ is

$$\frac{dy}{dx} = \frac{\partial y}{\partial x} + \frac{\partial y}{\partial z}\frac{dz}{dx} \tag{27}$$

Let us now see how the above equation is arrived at.

### *2.2.3. Partial and full derivatives through influence diagrams*

*Example 1:*

Consider the following relation again:

$$y = f(x, g(x)) \tag{28}$$

We wish to compute the full derivative of $y$ w.r.t. $x$.

We can rewrite the above equation as

$$z = g(x) \tag{29}$$
$$y = f(x, z) \tag{30}$$

The complete influence diagram is given below. This is identical to the figure of 2b below.

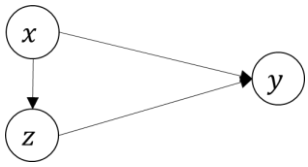


Figure 4: Influence diagram for Equation 28, represented using the decomposition of Equations 29 and 30.

In order to compute the full derivative of $y$ w.r.t $x$, we will consider the relations in the figure in steps. First, consider Equation 30. This has the following influence diagram.

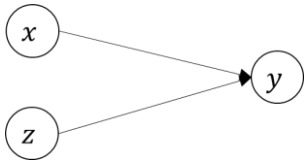


Figure 5: Influence diagram $y = f(x, z)$ (Equation 30)

This is just Figure 4 with the dependence of $z$ on $x$ masked out. Following Equation 23, the derivative rule is given by

$$\delta y = \frac{\partial y}{\partial x}\delta x + \frac{\partial y}{\partial z}\delta z \tag{31}$$

Note that all derivatives here are partial derivatives, since the *only* influences we use when computing the derivative rules are those from Figure 5.

Now, consider the remaining dependency. This is shown by Figure 6. This is the same as Figure 4 with the portion already considered in Figure 5 masked out. The derivative relation for this figure is given by

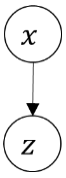


Figure 6: Influence diagram $z = g(x)$ (Equation 29)

$$\delta z = \frac{dz}{dx}\delta x \tag{32}$$

Combining Equation 32 with Equation 31, we get

$$\delta y = \left(\frac{\partial y}{\partial x} + \frac{\partial y}{\partial z}\frac{dz}{dx}\right)\delta x \tag{33}$$

From inspection of Equation 33, we come to the conclusion that the full derivative of $y$ with respect to $x$ is given the following equation

$$\frac{dy}{dx} = \frac{\partial y}{\partial x} + \frac{\partial y}{\partial z}\frac{dz}{dx} \tag{34}$$

■

*Example 2:*

Consider the following equation

$$y = f(g_1(x), g_2(x), \dots, g_K(x)) \tag{35}$$

Our objective is to compute the derivative of $y$ w.r.t. $x$. We can rewrite the equation as

$$z_i = g_i(x) \text{ for } i = 1 \dots K \tag{36}$$
$$y = f(z_1, z_2, \dots, z_K) \tag{37}$$

The influence diagram for this relation is given below

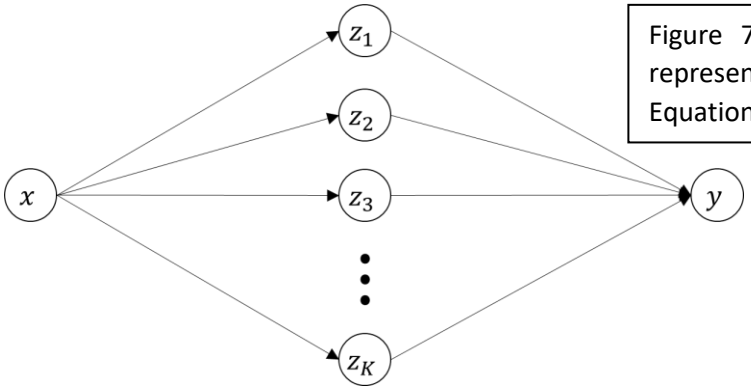


Figure 7: Influence diagram for Equation 35, represented using the decomposition of Equations 36 and 3.

To compute the full derivative of $y$ w.r.t. $x$ we can separate the figure into two parts as before:

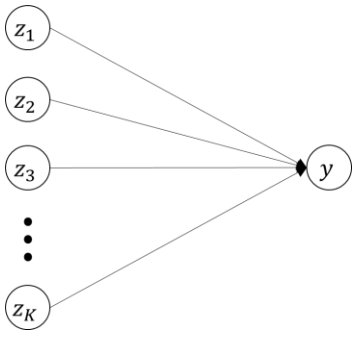


Figure 8: Influence diagram for $y = f(z_1, z_2, \dots, z_K)$

which gives us

$$\delta y = \frac{\partial y}{\partial z_1}\delta z_1 + \frac{\partial y}{\partial z_2}\delta z_2 + \cdots + \frac{\partial y}{\partial z_K}\delta z_K \tag{38}$$

and

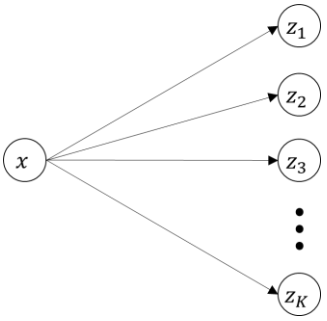


Figure 9: Influence diagram for $z_i = g_i(x)$ for $i = 1 \ldots K$

which gives us

$$\delta z_i = \frac{dz_i}{dx}\delta x \text{ for } i = 1 \ldots K \tag{39}$$

Combining Equation 38 with Equation 39, we get

$$\delta y = \left(\frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \frac{\partial y}{\partial z_2}\frac{dz_2}{dx} + \cdots + \frac{\partial y}{\partial z_K}\frac{dz_K}{dx}\right)\delta x \tag{40}$$

and the following equation for full derivative of $y$ with respect to $x$

$$\frac{dy}{dx} = \frac{\partial y}{\partial z_1}\frac{dz_1}{dx} + \frac{\partial y}{\partial z_2}\frac{dz_2}{dx} + \cdots + \frac{\partial y}{\partial z_K}\frac{dz_K}{dx} \tag{41}$$

■

*Example 3:*

Lets now consider a slightly more complex problem

$$y = f\big(x, g(x), h(x, g(x))\big) \tag{42}$$

The influence diagram for this relation is given by

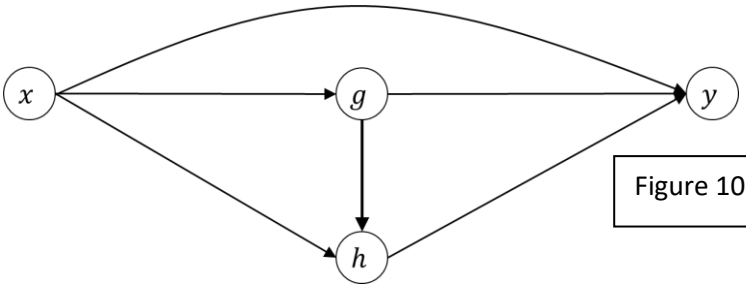


Figure 10: Influence diagram for Equation 42

As before, lets decompose the graph into the following three components.

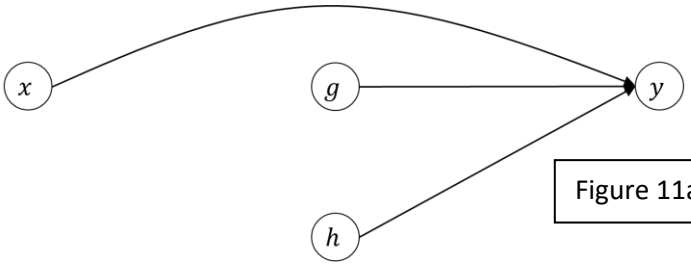


Figure 11a

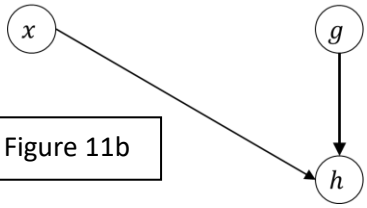


Figure 11b

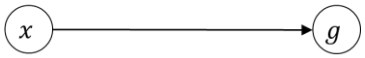


Figure 11c

It is important to note that the decomposition is not arbitrarily chosen. It follows the following rules.

a) Each subgraph consists of one *sink* node into which one or more arrows arrive, and all of the nodes that point into this node.
b) Every node with incoming edges has its own subgraph.
c) No edges are repeated across subgraphs.
d) The subgraphs, together, compose the entire original graph.

We can now write down the derivative rules for each of the subgraphs. From Figure 11a we get

$$\delta y = \frac{\partial y}{\partial x}\delta x + \frac{\partial y}{\partial g}\delta g + \frac{\partial y}{\partial h}\delta h \tag{43}$$

From Figure 11b we get

$$\delta h = \frac{\partial h}{\partial g}\delta g + \frac{\partial h}{\partial x}\delta x \tag{44}$$

From Figure 11c we get

$$\delta g = \frac{dg}{dx}\delta x \tag{45}$$

(Note that some of these have been written as partial derivatives, and others are full derivatives. We leave you to figure out why, from the figures).
Inserting Equations 44 and 45 in Equation 43, we get

$$\delta y = \frac{\partial y}{\partial x}\delta x + \frac{\partial y}{\partial g}\frac{dg}{dx}\delta x + \frac{\partial y}{\partial h}\left(\frac{\partial h}{\partial g}\frac{dg}{dx}\delta x + \frac{\partial h}{\partial x}\delta x\right) \tag{46}$$

giving us

$$\delta y = \left(\frac{\partial y}{\partial x} + \frac{\partial y}{\partial g}\frac{dg}{dx} + \frac{\partial y}{\partial h}\left(\frac{\partial h}{\partial g}\frac{dg}{dx} + \frac{\partial h}{\partial x}\right)\right)\delta x \tag{47}$$

and thus the derivative of $y$ with respect to $x$ as

$$\frac{dy}{dx} = \frac{\partial y}{\partial x} + \frac{\partial y}{\partial g}\frac{dg}{dx} + \frac{\partial y}{\partial h}\left(\frac{\partial h}{\partial g}\frac{dg}{dx} + \frac{\partial h}{\partial x}\right) \tag{48}$$