A vector function of a vector variable is a function $\mathbf{F}\colon \mathbb{R}^n \to \mathbb{R}^m$.

In practice, if $\langle x_1, \cdots x_n\rangle$ is the input,

$$\mathbf{F}(x_1, \cdots, x_n) = \big\langle F_1(x_1, \cdots, x_n), \cdots, F_m(x_1, \cdots, x_n)\big\rangle$$

where each $F_i(x_1, \cdots, x_n)$ is a multi-variable function of the sort in which we are currently interested.

Considering vector functions makes compositions easy to describe and makes the Chain Rule both easy to write down and easy to use.

Here is an example of a composition with vector functions and without vector functions.

If $\mathbf{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{G}\colon \mathbb{R}^m \to \mathbb{R}^p$ are given, then

$$(\mathbf{G} \circ \mathbf{F})(x_1, \cdots, x_n) = \mathbf{G}\big(F_1(x_1, \cdots, x_n), \cdots, F_m(x_1, \cdots, x_n)\big) =$$

$$\Big(G_1\big(F_1(x_1, \cdots, x_n), \cdots, F_m(x_1, \cdots, x_n)\big), \cdots, G_p\big(F_1(x_1, \cdots, x_n), \cdots, F_m(x_1, \cdots, x_n)\big)\Big)$$

If I write $\mathbf{x} = \langle x_1, \cdots, x_n\rangle$ then I can also write

$$(\mathbf{G} \circ \mathbf{F})(\mathbf{x}) = \mathbf{G}\big(\mathbf{F}(\mathbf{x})\big)$$

So rather than describing $(u^2+v^2-uw)^3+(u^3+v^3-vw)^2$ as the composition of $g(x, y) = x^3+y^2$ with $x(u, v) = u^2 + v^2 - uw$ and $y(u, v) = u^3 + v^3 - vw$ I can describe it as g composed with the vector function $\mathbf{F}(u, v, w) = \langle u^2 + v^2 - uw, u^3 + v^3 - vw\rangle$. Even if I only care about real-valued functions of a vector variable, as soon as I want to do composites, vector functions give me the same advantages that vectors have over lists of numbers.

The Chain Rule is a formula for $\dfrac{\partial\, G \circ \mathbf{F}}{\partial\, x_i}$ where $G\colon \mathbb{R}^m \to \mathbb{R}$ is a real-valued function of m variables and $\mathbf{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ is a vector function. Let us write $H(x_1, \cdots, x_n)$ for $G \circ \mathbf{F}$. In this notation, we are looking for a formula for $\dfrac{\partial\, H}{\partial\, x_i}$.

The Chain Rule

If $H(\mathbf{x}) = G\big(\mathbf{F}(\mathbf{x})\big)$ and if x_i is one of the input variables for $\mathbf{F}$, then

$$(CR) \qquad \frac{\partial H}{\partial x_i} = \nabla G\big(\mathbf{F}(\mathbf{x})\big) \bullet \frac{\partial \mathbf{F}}{\partial x_i}(\mathbf{x})$$

provided G and $\mathbf{F}$ are differentiable.

Let's start by talking about the $\dfrac{\partial \mathbf{F}}{\partial x_i}(\mathbf{x})$ piece. If we fix all the x_j's except for x_i, we get a curve in $\mathbb{R}^m$.

Examples.

- Let $\mathbf{F}(x, y, z) = \langle x^2 - y^2, xyz\rangle$. Fix x and z, say $x = 2$, $z = 3$.
  Then the curve in $\mathbb{R}^2$ is $\langle 4 - y^2, 6y\rangle$.
- Let $\mathbf{F}(x, y) = \langle x^2 - y^2, xy, x^3 - 3xy^2\rangle$. Fix $y = 2$.
  Then the curve in $\mathbb{R}^3$ is $\langle x^2 - 4, 2x, x^3 - 12x\rangle$.

From our work on curves we know how to write down a tangent vector to a curve:

$$\frac{\partial \mathbf{F}}{\partial x_i} = \left\langle \frac{\partial F_1}{\partial x_i}, \cdots, \frac{\partial F_m}{\partial x_i} \right\rangle$$

In the case we already studied in Chapter 13 we wrote $\mathbf{F}$ as $\mathbf{r}(t)$ and

$$\frac{\partial \mathbf{F}}{\partial t} = \mathbf{r}'(t)$$

As far as calculations go the general case is the same since only the variable x_i is allowed to move.

Example.

- Let $\mathbf{F}(x, y, z) = \langle x^2 - y^2, xyz\rangle$. Then
  $\dfrac{\partial \mathbf{F}}{\partial x} = \langle 2x, yz\rangle$, $\dfrac{\partial \mathbf{F}}{\partial y} = \langle -2y, xz\rangle$ and $\dfrac{\partial \mathbf{F}}{\partial z} = \langle 0, xy\rangle$.
- Let $\mathbf{F}(x, y) = \langle x^2 - y^2, xy, x^3 - 3xy^2\rangle$. Then
  $\dfrac{\partial \mathbf{F}}{\partial x} = \langle 2x, y, 3x^2 - 3y^2\rangle$ and $\dfrac{\partial \mathbf{F}}{\partial y} = \langle -2y, x, -6xy\rangle$.

Turn now to $G(y_1, \cdots, y_m)$ a multi-variable function. We will study the *gradient* intensely in the next few lectures, but for now we only need the definition.

$$\nabla G(y_1, \cdots, y_m) = \left\langle \frac{\partial G}{\partial y_1}, \cdots, \frac{\partial G}{\partial y_m} \right\rangle$$

Notice that ∇G is a vector function $\mathbb{R}^m \to \mathbb{R}^m$ since each $\dfrac{\partial G}{\partial y_i}$ is a function $\mathbb{R}^m \to \mathbb{R}^1$.

> *Notice that we write down the partial derivatives in the same order as we write down the variables.*

Example.

- Let $G(x, y) = x^2 - xy + \sin y$. Then $\nabla G(x, y) = \langle 2x - y, -x + \cos y \rangle$.
- Let $G(x, y, z) = x^2 - xy + z \sin y$. Then $\nabla G(x, y, z) = \langle 2x - y, -x + z \cos y, \sin y \rangle$.

Let us work through some examples.

One variable calculus. The function G is a function of one variable, as is the function $\mathbf{F}$. The gradient $\nabla G(x)$ is a 1-vector $G'(x)$. The tangent vector $\dfrac{\partial \mathbf{F}}{\partial x}(x)$ is the 1-vector $\mathbf{F}'(x)$. The dot product in this case is just the product and so

$$H'(x) = G'\big(\mathbf{F}(x)\big)\mathbf{F}'(x)$$

In English, to differentiate a composition, take the derivative of the outside function, plug in the inside function, and then multiply by the derivative of the inside function.

The multi-variable chain rule has a similar description.

$$\frac{\partial G\big(\mathbf{F}(\mathbf{x})\big)}{\partial x_i} = \nabla G\big(\mathbf{F}(\mathbf{x})\big) \bullet \frac{\partial \mathbf{F}}{\partial x_i}(\mathbf{x})$$

> *To find a partial derivative of a composition, take the gradient of the outside function, plug in the inside function and then take the dot product with the partial derivative of the inside function.*

Stewart, Case 1. $z = f(x, y)$, $x = g(t)$, $y = h(t)$ and $z(t)$ is the compoosition. Then the vector function is $\mathbf{r}(t) = \langle g(t), h(t)\rangle$ so $z(t) = f\big(\mathbf{r}(t)\big)$. Then $\dfrac{\partial \mathbf{r}}{\partial t} = \mathbf{r}'(t) = \left\langle \dfrac{dg}{dt}, \dfrac{dh}{dt} \right\rangle$, $\nabla f = \left\langle \dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y} \right\rangle$ and

$$\frac{dz}{dt} = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \bullet \left\langle \frac{dg}{dt}, \frac{dh}{dt} \right\rangle = \frac{\partial f}{\partial x}\frac{dg}{dt} + \frac{\partial f}{\partial y}\frac{dh}{dt}$$

Writing the formula this way hides the fact that you must plug $x = g(t)$, $y = h(t)$ into the formulas for $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$.
So if $f(x, y) = x^2 + y^3$, $x(t) = t^2 - 1$, $y(t) = t^3 + t$, the correct calculation is

- $\mathbf{r}'(t) = \langle 2t, 3t^2 + 1\rangle$,
- $\nabla f = \langle 2x, 3y^2\rangle = \langle 2(t^2 - 1), 3(t^3 + t)^2\rangle$

and

$$\frac{dz}{dt} = \langle 2(t^2 - 1), 3(t^3 + t)^3\rangle \bullet \langle 2t, 3t^2 + 1\rangle = 4t(t^2 - 1) + 3(3t^2 + 1)(t^3 + t)^3 = \cdots$$

The correct answer is NOT $\dfrac{dz}{dt} = \langle 2x, 3y^2\rangle \bullet \langle 2t, 3t^2 + 1\rangle = 4xt + 3y^2(3t^2 + 1)$. A helpful clue is to remember that after the composition, z is a function of t and therefore its derivative must also be a function of t, NOT t, x and y. As we will see below, in certain cases it is possible to work with this mixed form to reduce the work required.

Stewart, Case 2. $z = f(x, y)$, $x = g(s, t)$, $y = h(s, t)$ and $z(s, t)$ is the composition. We can compute $\dfrac{\partial z}{\partial s}$ or $\dfrac{\partial z}{\partial t}$. First compute $\nabla f = \left\langle \dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y} \right\rangle$ (and remember to plug in to write x and y's and functions of s and t). The vector function is $\mathbf{P}(s, t) = \langle g(s, t), h(s, t)\rangle$. There are two curve parts, one where t is fixed one where s is fixed: $\left\langle \dfrac{\partial g}{\partial s}, \dfrac{\partial h}{\partial s} \right\rangle$ and $\left\langle \dfrac{\partial g}{\partial t}, \dfrac{\partial h}{\partial t} \right\rangle$.
Unraveling the dot product

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} \quad \text{and} \quad \frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$$

So this time, if $f(x, y) = x^2 + y^3$, $x(s, t) = t^2 - s^2$, $y(s, t) = st$, the correct calculation is

- $\dfrac{\partial \mathbf{P}}{\partial s} = \langle -2s, t\rangle$ and $\dfrac{\partial \mathbf{P}}{\partial t} = \langle 2t, s\rangle$
- $\nabla f = \langle 2x, 3y^2\rangle = \langle 2(t^2 - s^2), 3s^2t^2\rangle$

and

$$\begin{aligned} \frac{\partial z}{\partial s} &= \langle 2(t^2 - s^2), 3s^2t^2 \rangle \bullet \langle -2s, t \rangle = -4s(t^2 - s^2) + 3s^2t^3 = \cdots \\ \frac{\partial z}{\partial t} &= \langle 2(t^2 - s^2), 3s^2t^2 \rangle \bullet \langle 2t, s \rangle = 4t(t^2 - s^2) + 3s^3t^2 = \cdots \end{aligned}$$

Note the partials of the composition should be functions of s and t only and they are.

Notice that the gradient is the same in both cases so you only have to compute it once. Looking at the general formula (CR) you see that this is always the case. No matter how many different partials of the composition you need to compute, the first vector in the dot product is always the same, the gradient with the vector function plugged in.

Stewart, Example 3. $z(x,y) = e^x \sin y$, $x = st^2$, $y = s^2 t$. Then $\mathbf{P}(s,t) = \langle st^2, s^2 t\rangle$.

- $\nabla z = \langle e^x \sin y, e^x \cos y\rangle = \langle e^{st^2} \sin(s^2 t), e^{st^2} \cos(s^2 t)\rangle$.
- $\dfrac{\partial \mathbf{P}}{\partial s} = \langle t^2, 2st\rangle$
- $\dfrac{\partial \mathbf{P}}{\partial t} = \langle 2st, s^2\rangle$

$$\begin{aligned} \frac{\partial z}{\partial s} &= \langle e^{st^2} \sin(s^2 t), e^{st^2} \cos(s^2 t)\rangle \bullet \langle t^2, 2st\rangle = t^2 e^{st^2} \sin(s^2 t) + 2ste^{st^2} \cos(s^2 t) \\ \frac{\partial z}{\partial t} &= \langle e^{st^2} \sin(s^2 t), e^{st^2} \cos(s^2 t)\rangle \bullet \langle 2st, s^2\rangle = 2ste^{st^2} \sin(s^2 t) + s^2 e^{st^2} \cos(s^2 t) \end{aligned}$$

Notice that no *tree diagrams* are need to keep everything straight. The dot product does that for you.

Stewart, Example 4. $w = f(x,y,z,t)$, $x = x(u,v)$, $y = y(u,v)$, $z = z(u,v)$ and $t = t(u,v)$ so the composition is $w(u,v)$. To compute $\dfrac{\partial w}{\partial u}$ and $\dfrac{\partial w}{\partial v}$ first write down the vector function, say $\mathbf{P}(u,v) = \langle x(u,v), y(u,v), z(u,v), t(u,v)\rangle$. Then $w(u,v) = f \circ \mathbf{P}$.

- $\nabla f = \left\langle \dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}, \dfrac{\partial f}{\partial z}, \dfrac{\partial f}{\partial t} \right\rangle$ (and of course you should think of ∇f as a function of u and v).
- $\dfrac{\partial \mathbf{P}}{\partial u} = \left\langle \dfrac{\partial x}{\partial u}, \dfrac{\partial y}{\partial u}, \dfrac{\partial z}{\partial u}, \dfrac{\partial t}{\partial u} \right\rangle$ and $\dfrac{\partial \mathbf{P}}{\partial v} = \left\langle \dfrac{\partial x}{\partial v}, \dfrac{\partial y}{\partial v}, \dfrac{\partial z}{\partial v}, \dfrac{\partial t}{\partial v} \right\rangle$.

$$\begin{aligned} \frac{\partial w}{\partial u} &= \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}, \frac{\partial f}{\partial t} \right\rangle \bullet \left\langle \frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u}, \frac{\partial t}{\partial u} \right\rangle \\ \frac{\partial w}{\partial v} &= \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}, \frac{\partial f}{\partial t} \right\rangle \bullet \left\langle \frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v}, \frac{\partial t}{\partial v} \right\rangle \end{aligned}$$

which you can easily verify is the answer in Stewart.

Stewart, Example 5. $u(x, y, z) = x^4y + y^2z^3$, $x = rse^t$, $y = rs^2e^{-t}$ $z = r^2s\sin t$. Find the value of $\dfrac{\partial u}{\partial s}$ when $r = 2$, $s = 1$ and $t = 0$. The vector function is $\mathbf{P}(r, s, t) = \langle rse^t, rs^2e^{-t}, r^2s\sin t\rangle$. The point in xyz space is $(2, 2, 0)$. The fact that we only want the partial at one point allows a simplification in the calculation as follows.

- $\nabla u = \langle 4x^3y, x^4 + 2yz^3, 3y^2z^2\rangle$. Normally I would next substitute for r, s and t but since I only want the value of ∇u when $r = 2$, $s = 1$ and $t = 0$ I can just plug in $x = 2$,$y = 2$ and $z = 0$ right now. $\nabla u = \langle 64, 16, 0\rangle$.
- $\dfrac{\partial \mathbf{P}}{\partial s} = \langle re^t, 2rse^{-t}, r^2\sin t\rangle$, which at $r = 2$, $s = 1$ and $t = 0$ is $\dfrac{\partial \mathbf{P}}{\partial s} = \langle 2, 4, 0\rangle$

Hence

$$\left.\frac{\partial u}{\partial s}\right|_{(2,1,0)} = \langle 64, 16, 0\rangle \bullet \langle 2, 4, 0\rangle = 128 + 64 = 192$$

To get the other two partials at the point, compute $\left.\dfrac{\partial \mathbf{P}}{\partial t}\right|_{(2,1,0)} = \langle rse^t, -rs^2e^{-t}, r^2s\cos t\rangle\Big|_{(2,1,0)} = \langle 2, -2, 4\rangle$ so

$$\left.\frac{\partial u}{\partial t}\right|_{(2,1,0)} = \langle 64, 16, 0\rangle \bullet \langle 2, -2, 4\rangle = 128 - 32 = 96$$

and $\left.\dfrac{\partial \mathbf{P}}{\partial r}\right|_{(2,1,0)} = \langle se^t, s^2e^{-t}, 2rs\sin t\rangle\Big|_{(2,1,0)} = \langle 1, 1, 0\rangle$ so

$$\left.\frac{\partial u}{\partial r}\right|_{(2,1,0)} = \langle 64, 16, 0\rangle \bullet \langle 1, 1, 0\rangle = 64 + 16 = 90$$

Stewart at one point in his calculation writes

$$(4x^3y)(re^t) + (x^4 + 2yz^3)(2rse^{-t}) + (3y^2z^2)(r^2\sin t)$$

mixing the two sets of variables, which is dangerous. But having written this down he still needs to substitute for x, y and z in terms of r, s and t, but since Stewart is at the point $(2, 1, 0)$ he substitutes into the original formulas for x y and z and gets $x = y = 2$, $z = 0$ and then plugs in which is OK.

Higher order partials. Of course the next issue is computing a second order partial of a composition. And then a third order partial and so on.

If $H = G \circ \mathbf{F}$ then we want to compute something like

$$\frac{\partial^2 H}{\partial x_i \, \partial x_j} = \frac{\partial \left(\dfrac{\partial H}{\partial x_j} \right)}{\partial x_i}$$

We definitely want both the case where $x_i \neq x_j$ and where $x_i = x_j$. The next part if grayed out because, while the formulas are correct the calculations can be simplified in terms of what you need to remember. In particular, as promised, to compute a higher derivative you simply differentiate with respect to the first variable, then differentiate that answer with respect to the second variable, and so on. Keep things written in terms of dot products and things remain as simple as possible. An example is provided after the grayed out material.

By the Chain Rule

$$\frac{\partial H(\mathbf{x})}{\partial x_j} = \nabla G\big(\mathbf{F}(\mathbf{x})\big) \bullet \frac{\partial \mathbf{F}}{\partial x_j}(\mathbf{x})$$

We now need to find $\dfrac{\partial \left(\dfrac{\partial H}{\partial x_j} \right)}{\partial x_i}$. By the product rule,

$$(C^2R) \qquad \frac{\partial \left(\dfrac{\partial H}{\partial x_j} \right)}{\partial x_i}(\mathbf{x}) = \frac{\partial \nabla G\big(\mathbf{F}(\mathbf{x})\big)}{\partial x_i} \bullet \frac{\partial \mathbf{F}}{\partial x_j}(\mathbf{x}) + \nabla G\big(\mathbf{F}(\mathbf{x})\big) \bullet \frac{\partial^2 \mathbf{F}}{\partial x_i \, \partial x_j}(\mathbf{x})$$

There are four terms in this formula. The gradient $\nabla G\big(\mathbf{F}(\mathbf{x})\big)$ and $\dfrac{\partial \mathbf{F}}{\partial x_j}(\mathbf{x})$ are just the same as you computed in for the Chain Rule. The term $\dfrac{\partial^2 \mathbf{F}}{\partial x_i \, \partial x_j}(\mathbf{x})$ is computed by taking the vector $\dfrac{\partial \mathbf{F}}{\partial x_j}(\mathbf{x})$ and differentiating it with respect to x_i. In practice this is straightforward.

The $\dfrac{\partial \nabla G\big(\mathbf{F}(\mathbf{x})\big)}{\partial x_i}$ term is also the partial derivative of a vector function. The k^{th} component in ∇G is $\dfrac{\partial G}{\partial x_k}$ composed with $\mathbf{F}$ so the k^{th} term in $\dfrac{\partial \nabla G\big(\mathbf{F}(\mathbf{x})\big)}{\partial x_i}$ is the vector

$$\left\langle \frac{\partial^2 G}{\partial x_1 \, \partial x_k}, \cdots, \frac{\partial^2 G}{\partial x_m \, \partial x_k} \right\rangle \bullet \frac{\partial \mathbf{F}}{\partial x_i} = \nabla \left(\frac{\partial G}{\partial x_k} \right) \bullet \frac{\partial \mathbf{F}}{\partial x_i}$$

Hence

$$\frac{\partial \nabla G\big(\mathbf{F}(\mathbf{x})\big)}{\partial x_i} = \left\langle \nabla \left(\frac{\partial G}{\partial x_1} \right) \bullet \frac{\partial \mathbf{F}}{\partial x_i}, \cdots, \nabla \left(\frac{\partial G}{\partial x_m} \right) \bullet \frac{\partial \mathbf{F}}{\partial x_i} \right\rangle$$

and

$$\frac{\partial^2 H}{\partial x_i \, \partial x_j} = \left\langle \nabla \left(\frac{\partial G}{\partial x_1} \right) \bullet \frac{\partial \mathbf{F}}{\partial x_i}, \cdots, \nabla \left(\frac{\partial G}{\partial x_m} \right) \bullet \frac{\partial \mathbf{F}}{\partial x_i} \right\rangle \bullet \frac{\partial \mathbf{F}}{\partial x_j}(\mathbf{x}) + \nabla G\big(\mathbf{F}(\mathbf{x})\big) \bullet \frac{\partial^2 \mathbf{F}}{\partial x_i \, \partial x_j}(\mathbf{x})$$

Especially if you are willing to use both sets of variables in your answer, the calculations are not too bad.

Stewart Example 7. Given an unknown function $z = f(x, y)$ and $x = r^2 + s^2$, $y = 2rs$ find $\frac{\partial^2 f}{\partial r^2}$. The vector function is $\mathbf{P} = \langle r^2 + s^2, 2rs\rangle$. The vector function $\nabla f = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle$ is basically unknown, but $\frac{\partial \mathbf{P}}{\partial r} = \langle 2r, 2s\rangle$. In the formula for the second partial we will need $\frac{\partial^2 \mathbf{P}}{\partial r^2} = \langle 2, 0\rangle$ and $\nabla\left(\frac{\partial f}{\partial x}\right) = \left\langle \frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial y\,\partial x} \right\rangle$, $\nabla\left(\frac{\partial f}{\partial y}\right) = \left\langle \frac{\partial^2 f}{\partial x\,\partial y}, \frac{\partial^2 f}{\partial y^2} \right\rangle$.

Let us assume as Stewart does that mixed partials are equal. Then

$$\left\langle \nabla\left(\frac{\partial f}{\partial x}\right) \bullet \frac{\partial \mathbf{P}}{\partial r}, \nabla\left(\frac{\partial f}{\partial y}\right) \bullet \frac{\partial \mathbf{P}}{\partial r} \right\rangle = \left\langle 2r\frac{\partial^2 f}{\partial x^2} + 2s\frac{\partial^2 f}{\partial x\,\partial y}, 2r\frac{\partial^2 f}{\partial x\,\partial y} + 2s\frac{\partial^2 f}{\partial y^2} \right\rangle$$

and

$$\frac{\partial^2 f}{\partial r^2} = \left\langle 2r\frac{\partial^2 f}{\partial x^2} + 2s\frac{\partial^2 f}{\partial x\,\partial y}, 2r\frac{\partial^2 f}{\partial x\,\partial y} + 2s\frac{\partial^2 f}{\partial y^2} \right\rangle \bullet \langle 2r, 2s\rangle + \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \bullet \langle 2, 0\rangle$$

which is the formula in Stewart.

The other two second order partials can be calculated as well. We will need $\frac{\partial \mathbf{P}}{\partial s} = \langle 2s, 2r\rangle$, $\frac{\partial^2 \mathbf{P}}{\partial r\,\partial s} = \langle 0, 2\rangle$ and $\frac{\partial^2 \mathbf{P}}{\partial s^2} = \langle 2, 0\rangle$. Then

$$\left\langle \nabla\left(\frac{\partial f}{\partial x}\right) \bullet \frac{\partial \mathbf{P}}{\partial s}, \nabla\left(\frac{\partial f}{\partial y}\right) \bullet \frac{\partial \mathbf{P}}{\partial s} \right\rangle = \left\langle 2s\frac{\partial^2 f}{\partial x^2} + 2r\frac{\partial^2 f}{\partial x\,\partial y}, 2s\frac{\partial^2 f}{\partial x\,\partial y} + 2r\frac{\partial^2 f}{\partial y^2} \right\rangle$$

and

$$\frac{\partial^2 f}{\partial s^2} = \left\langle 2s\frac{\partial^2 f}{\partial x^2} + 2r\frac{\partial^2 f}{\partial x\,\partial y}, 2s\frac{\partial^2 f}{\partial x\,\partial y} + 2r\frac{\partial^2 f}{\partial y^2} \right\rangle \bullet \langle 2s, 2r\rangle + \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \bullet \langle 2, 0\rangle$$

and

$$\frac{\partial^2 f}{\partial r\,\partial s} = \left\langle 2s\frac{\partial^2 f}{\partial x^2} + 2r\frac{\partial^2 f}{\partial x\,\partial y}, 2s\frac{\partial^2 f}{\partial x\,\partial y} + 2r\frac{\partial^2 f}{\partial y^2} \right\rangle \bullet \langle 2r, 2s\rangle + \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \bullet \langle 0, 2\rangle$$

$\frac{\partial^2 f}{\partial r\,\partial s}$ can also be computed using

$$\left\langle \nabla\left(\frac{\partial f}{\partial x}\right) \bullet \frac{\partial \mathbf{P}}{\partial r}, \nabla\left(\frac{\partial f}{\partial y}\right) \bullet \frac{\partial \mathbf{P}}{\partial r} \right\rangle \bullet \frac{\partial \mathbf{P}}{\partial s} = \left\langle 2r\frac{\partial^2 f}{\partial x^2} + 2s\frac{\partial^2 f}{\partial x\,\partial y}, 2r\frac{\partial^2 f}{\partial x\,\partial y} + 2s\frac{\partial^2 f}{\partial y^2} \right\rangle \bullet \langle 2s, 2r\rangle$$

Let us redo this last example without trying to plug into formula (C^2R). In particular, without having to memorize the formula (C^2R)!

Given an unknown function $z = f(x, y)$ and $x = r^2 + s^2$, $y = 2rs$ find $\dfrac{\partial^2 f}{\partial r^2}$. The vector function is $\mathbf{P} = \langle r^2 + s^2, 2rs\rangle$. The vector function $\nabla f = \left\langle \dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y} \right\rangle$ is basically unknown, but $\dfrac{\partial \mathbf{P}}{\partial r} = \langle 2r, 2s\rangle$.

The Chain Rule says

$$\frac{\partial f}{\partial r} = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \bullet \langle 2r, 2s\rangle$$

A Product Rule says

$$\begin{aligned} \frac{\partial^2 f}{\partial^2 r} &= \frac{\partial}{\partial r}\left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \bullet \langle 2r, 2s\rangle + \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \bullet \langle 2, 0\rangle \\ (*) \qquad &= \left\langle \frac{\partial}{\partial r}\frac{\partial f}{\partial x}, \frac{\partial}{\partial r}\frac{\partial f}{\partial y} \right\rangle \bullet \langle 2r, 2s\rangle + \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \bullet \langle 2, 0\rangle \end{aligned}$$

Apply the Chain Rule again

$$\begin{aligned} \frac{\partial}{\partial r}\frac{\partial f}{\partial x} &= \nabla\left(\frac{\partial f}{\partial x}\right) \bullet \frac{\partial \mathbf{P}}{\partial r} = \left\langle \frac{\partial^2 f}{\partial^2 x}, \frac{\partial^2 f}{\partial y \partial x} \right\rangle \bullet \langle 2r, 2s\rangle = 2r\frac{\partial^2 f}{\partial^2 x} + 2s\frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial}{\partial r}\frac{\partial f}{\partial y} &= \nabla\left(\frac{\partial f}{\partial y}\right) \bullet \frac{\partial \mathbf{P}}{\partial r} = \left\langle \frac{\partial^2 f}{\partial x \partial y}, \frac{\partial^2 f}{\partial^2 y} \right\rangle \bullet \langle 2r, 2s\rangle = 2r\frac{\partial^2 f}{\partial x \partial y} + 2s\frac{\partial^2 f}{\partial^2 y} \end{aligned}$$

so

$$\left\langle \frac{\partial}{\partial r}\frac{\partial f}{\partial x}, \frac{\partial}{\partial r}\frac{\partial f}{\partial y} \right\rangle = \left\langle 2r\frac{\partial^2 f}{\partial^2 x} + 2s\frac{\partial^2 f}{\partial y \partial x}, 2r\frac{\partial^2 f}{\partial x \partial y} + 2s\frac{\partial^2 f}{\partial^2 y} \right\rangle$$

and plugging into (∗) gives the same result as on the last page.

$$\frac{\partial^2 f}{\partial^2 r} = \left\langle 2r\frac{\partial^2 f}{\partial^2 x} + 2s\frac{\partial^2 f}{\partial y \partial x}, 2r\frac{\partial^2 f}{\partial x \partial y} + 2s\frac{\partial^2 f}{\partial^2 y} \right\rangle \bullet \langle 2r, 2s\rangle + \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right\rangle \bullet \langle 2, 0\rangle$$

If you need a third order partial, just differentiate this last equation. This time you will use the Sum Rule three times, the Dot Product Rule twice, the Scalar Product Rule four times and the Chain Rule six times. Nobody said it was going to be short!

But, no matter how complicated it gets, you are always just computing the derivative of sums, dot products, vector functions and applying the Chain Rule with the same vector function $\mathbf{F}$ when computing a higher order partial of $G\big(\mathbf{F}(\mathbf{x})\big)$.

1. Implicit differentiation

The Chain Rule in more than one variable is not so important for doing calculations since you can always do the substitutions and then compute the partial(s) you want directly. In one variable of course this is not the case. Just because you know how to differentiate $\sin x$, without the Chain Rule, you have no idea what to do with $\sin(x^2)$.

In more than one variable however, the Chain Rule is behind many of the deepest applications of multi-variable calculus. One example is implicit differentiation.

Let $F(x_1, \cdots, x_n, w)$ be a function of $n+1$ variables and suppose that there is a function of n variables, $w = w(x_1, \cdots, x_n)$ such that $F\big(x_1, \cdots, x_n, w(x_1, \cdots, x_n)\big) = c$ for some constant c. Then we say F defines w implicitly as a function of $x_1, \ldots, x_n$.

Suppose $(a_1, \cdots, a_n, w_0)$ is some point with F of that point equal to c. Then the Chain Rule allows us to compute the partials of w with respect to x_i at the point $(a_1, \cdots, a_n)$ without knowing the function w explicitly. To derive the formula note let $\mathbf{P}(x_1, \cdots, x_n) = \big\langle x_1, \cdots, x_n, w(x_1 \cdots, x_n)\big\rangle$ so $F \circ \mathbf{P} = c$.

Compute $\dfrac{\partial F \circ \mathbf{P}}{\partial x_i}$ two ways. Since it is a constant function $\dfrac{\partial F \circ \mathbf{P}}{\partial x_i} = 0$. Applying the Chain Rule, it is also $\nabla F\big(\mathbf{P}(\mathbf{x})\big) \bullet \dfrac{\partial \mathbf{P}}{\partial x_i}$. But $\dfrac{\partial \mathbf{P}}{\partial x_i}$ has all its components 0 except for the i^{th} which is 1 and the $(n+1)^{\text{st}}$ which is $\dfrac{\partial w}{\partial x_i}$. Hence $\dfrac{\partial F}{\partial x_i} + \dfrac{\partial F}{\partial w}\dfrac{\partial w}{\partial x_i} = 0$ or

$$\text{(ID)} \qquad \frac{\partial w}{\partial x_i} = -\frac{\dfrac{\partial F}{\partial x_i}}{\dfrac{\partial F}{\partial w}}$$

which is defined as long as $\dfrac{\partial F}{\partial w}$ is not zero at the point.

The Implicit Function Theorem says that if F has continuous partials in a neighborhood of $\mathbf{a} = \langle a_1, \cdots, a_n, w_0\rangle$ and if $\left.\dfrac{\partial F}{\partial w}\right|_{\mathbf{a}} \neq 0$, then there is a unique differentiable function $w = w(x_1, \cdots, x_n)$ defined in a neighborhood of $\mathbf{a}$ whose partials are given by (ID) and such that $w_0 = w(a_1, \cdots, a_n)$.

Example. Let $F(x, y, z) = x^2 - 3xy - xz^3 - y^2z^2$ and let $\mathbf{a} = (2, 3, -2)$. Then $F(\mathbf{a}) = -34$. Then the equation $F(x, y, z) = -34$ defines z implicitly as a function of x and y in a neighborhood of $\mathbf{a}$. Compute $\dfrac{\partial F}{\partial x} = 2x - 3y - z^3$, $\dfrac{\partial F}{\partial y} = -3x - 2y^2z$ and $\dfrac{\partial F}{\partial z} = -3xz^2 - 2y^2z$ and at $\mathbf{a}$, $\dfrac{\partial F}{\partial x} = 3$, $\dfrac{\partial F}{\partial y} = -30$ and $\dfrac{\partial F}{\partial z} = 12$. Since $\dfrac{\partial F}{\partial z} \neq 0$, there really is some function $z(x, y)$ such that $z(2, 3) = -2$ and $F\big(x, y, z(x, y)\big) = -34$ and z is guaranteed to be defined and differential in some (undetermined) neighborhood of $(2,3)$. Moreover

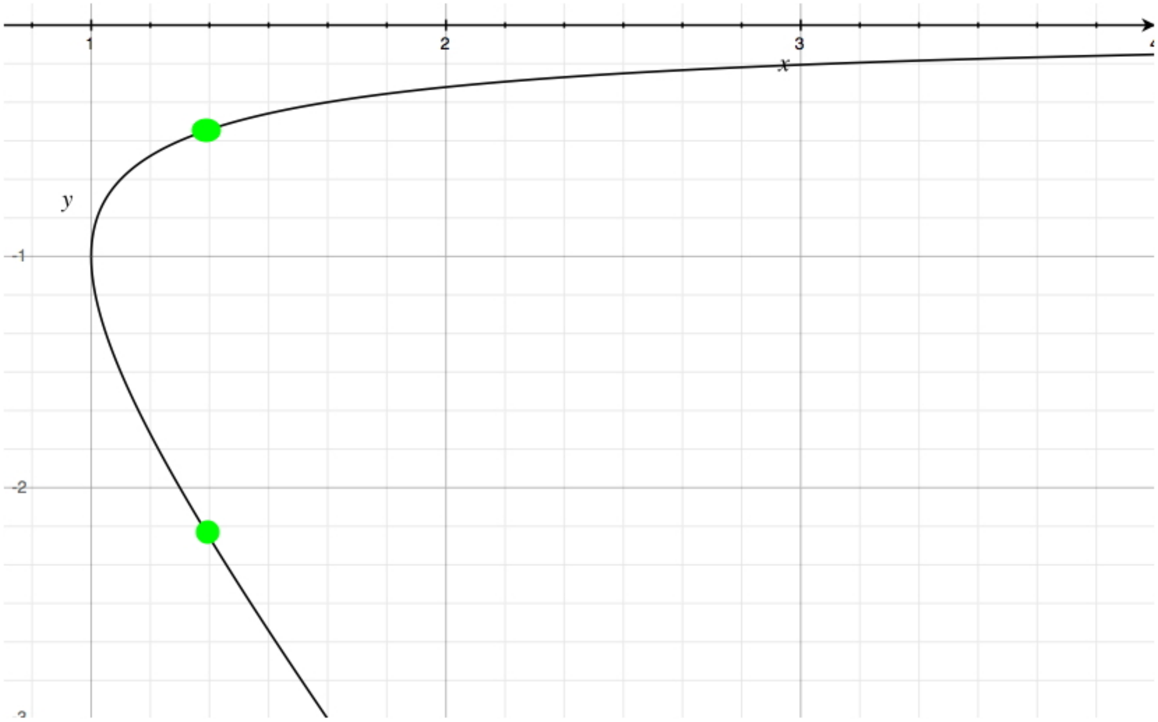
$$\frac{\partial z}{\partial x} = -\frac{2x - 3y - z^3}{-3xz^2 - 2y^2 z} = \frac{2x - 3y - z^3}{3xz^2 + 2y^2 z}$$

$$\frac{\partial z}{\partial y} = -\frac{-3x - 2yz^2}{-3xz^2 - 2y^2 z} = \frac{-3x - 2yz^2}{3xz^2 + 2y^2 z}$$

in that neighborhood.

Example. Here is a piece of the curve

$$y^2 + 2xy + 1 = 0$$



which is part of the level curve of $F(x, y) = y^2 + 2xy + 1$ with constant 0. At either green point on the curve, F clearly defines y as a function of x near the point. In other words, the curve passes the vertical line test if we restrict to a small interval around either green dot. Equivalently, a small piece of the curve near one of these points is the graph of a function. In this case, using the quadratic formula one can even write down a formula. Near the top green point $y = -x + \sqrt{x^2 - 1}$.

Compute $\dfrac{\partial F}{\partial x} = 2y$ and $\dfrac{\partial F}{\partial y} = 2y + 2x$. Hence

$$\frac{dy}{dx} = \frac{\partial y}{\partial x} = -\frac{2y}{2y + 2x} = \frac{-y}{y + x}$$

At the top green point $y + x = \sqrt{x^2 - 1} > 0$ and the curve is increasing. At the bottom green point $y + x = -\sqrt{x^2 - 1} < 0$ and the curve is decreasing. The point $(1, -1)$ is on the curve, but there $\dfrac{\partial F}{\partial y} = 0$ and the Implicit Function Theorem says nothing. Looking at the graph we see there is no implicitly defined function around the point $(1, -1)$.

It is also possible to compute higher derivatives. Recall $\frac{d^2 y}{dx^2}$ tells about concavity. Well,

$$\frac{d\,y}{d\,x} = \frac{\partial\,y}{\partial\,x} = -\frac{2y}{2y+2x} = \frac{-y}{y+x}$$

and

$$\frac{d^2y}{d\,x^2} = \frac{\partial^2 y}{\partial\,x^2} = \frac{\frac{\partial\,(-y)}{\partial\,x}(y+x) - (-y)\frac{\partial\,(y+x)}{\partial\,x}}{(y+x)^2} = \frac{-\frac{\partial\,y}{\partial\,x}(y+x) + y\left(\frac{\partial\,y}{\partial\,x}+1\right)}{(y+x)^2} =$$

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

so

$$\frac{d^2y}{d\,x^2} = \frac{y(y+2x)}{(y+x)^3}$$

For concavity you want to know if $\frac{d^2y}{d\,x^2}$ is positive or negative or zero for potential inflection points. Hence we need to determine the signs of y, $y+x$ and $y+2x$. At the two green points $x \approx 1.3$. For the top point $y \approx -0.5$ and for the lower point $y \approx -2.16$. Hence y is negative for both points. The quantity $y+x$ is positive for the top point and negative for the bottom. The quantity $y+2x$ is positive for both points. Hence the curve is concave down at the top point and concave up at the bottom point.

Likewise for the three-variable example above. Recall

$$\frac{\partial\,z}{\partial\,x} = \frac{2x-3y-z^3}{3xz^2+2y^2z}$$

$$\frac{\partial\,z}{\partial\,y} = \frac{-3x-2yz^2}{3xz^2+2y^2z}$$

Then

$$\frac{\partial^2 z}{\partial\,y\,\partial\,x} = \frac{\partial\left(\frac{\partial\,z}{\partial\,x}\right)}{\partial\,y} = \frac{\partial\,\frac{2x-3y-z^3}{3xz^2+2y^2z}}{\partial\,y} =$$

$$\frac{\frac{\partial\,2x-3y-z^3}{\partial\,y}(3xz^2+2y^2z) - (2x-3y-z^3)\frac{\partial\,3xz^2+2y^2z}{\partial\,y}}{(3xz^2+2y^2z)^2} =$$

$$\frac{\left(-3-3z^2\frac{\partial\,z}{\partial\,y}\right)(3xz^2+2y^2z) - (2x-3y-z^3)\left(\frac{\partial\,3xz^2+2y^2z}{\partial\,y}\right)}{(3xz^2+2y^2z)^2} =$$

$$\frac{\left(-3-3z^2\frac{\partial\,z}{\partial\,y}\right)(3xz^2+2y^2z) - (2x-3y-z^3)\left(0z^2+6xz\frac{\partial\,z}{\partial\,y}+4yz+4y^2\frac{\partial\,z}{\partial\,y}\right)}{(3xz^2+2y^2z)^2} = \cdots$$

The matrix version of the Chain Rule. If you picked up some matrix theory elsewhere in your life the ultimate version of the Chain Rule is the easiest of all. Start with a function $\mathbf{F}\colon \mathbb{R}^n \to \mathbb{R}^m$. If $\mathbf{F}(x_1, \cdots, x_n) = \langle F_1(x_1, \cdots, x_n), \cdots, F_m(x_1, \cdots, x_n)\rangle$ write down the $m \times n$ matrix

$$\mathbf{D}(\mathbf{F})(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial F_1}{\partial x_1} & \dfrac{\partial F_1}{\partial x_2} & \cdots & \dfrac{\partial F_1}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \dfrac{\partial F_m}{\partial x_1} & \dfrac{\partial F_m}{\partial x_2} & \cdots & \dfrac{\partial F_m}{\partial x_n} \end{bmatrix}$$

If $\mathbf{G}\colon \mathbb{R}^m \to \mathbb{R}^p$ then $\mathbf{G} \circ \mathbf{F}\colon \mathbb{R}^n \to \mathbb{R}^p$ is defined and

$$\mathbf{D}\big(\mathbf{G} \circ \mathbf{F}\big)(\mathbf{x}) = \Big(\mathbf{D}(\mathbf{G})\big(\mathbf{F}(\mathbf{x})\big)\Big)\Big(\mathbf{D}(\mathbf{F})(\mathbf{x})\Big)$$

where the right hand side is the product of the two matrices.

The formulas in the first part of this note follow from expressing the matrix multiplications required in terms of the dot product.