Chapter 9

theory of differentiation

In the last chapter I began by announcing I would apply the central theorems of this course to solve interesting applied problems. If you remembered that I said that you may be a bit perplexed after completing the preceding chapter. Where did we use these theorems? It would seem we mostly just differentiated and pulled a magic λ from the thin air. Where did we use the inverse or implicit mapping theorems? It's subtle. These theorems go to the existence of a mapping, or the solution of a system of equations. Often we do not even care about finding the inverse or solving the system. The mere existence justifies other calculations we do make explicit. In this chapter I hope to state the inverse and implicit function theorems carefully. I leave the complete proofs for Edward's text, we will just discuss portions of the proof. In particular, I think it's worthwhile to discuss Newton's method and the various generalizations which reside at the heart of Edward's proof. In contrast, I will take it easy on the analysis. The arguments given in Edward's generalize easily to the infinite dimensional case. I do think there are easy arguments but part of his game-plan is set-up the variational calculus chapter which is necessarily infinite-dimensional. Finally, I conclude this chapter by examining a few examples of constrained partial differentiation.

9.1 Newton's method for solving the insolvable

I'll begin with a quick review of Newton's method for functions.

Problem: given a function $f : \mathbb{R} \to \mathbb{R}$ which is continuously differentiable on $[a, b]$ and $f(a) < 0 < f(b)$ with $f'(x) > 0$ for each $x \in [a, b]$ **how** can we find the solution to $f(x) = 0$ w.r.t. the interval $[a, b]$?

Solution: Newton's Method. In a nutshell, the idea is to guess some point in $x_o \in [a, b]$ and then replace the function with the tangent line to $(x_o, f(x_o))$. Then we can easily calculate the zero of the tangent line through elementary algebra.

$$y = L_f^{x_o}(x) = f(x_o) + f'(x_o)(x - x_o) = 0 \qquad \Rightarrow \qquad x = x_o - \frac{f(x_o)}{f'(x_o)}$$

Now, this is just the first approximation, we can apply the idea again to our new guess $x_1 = x$; that is define $x_1 = x_o - \frac{f(x_o)}{f'(x_o)}$ and think of x_1 as our new "x_o". The zero of the tangent line to $(x_1, f(x_1))$ is called x_2 and we can calculate,

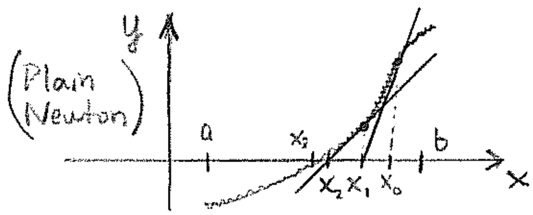
$$y = L_f^{x_1}(x) = f(x_1) + f'(x_1)(x - x_1) = 0 \qquad \Rightarrow \qquad \boxed{x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}}$$

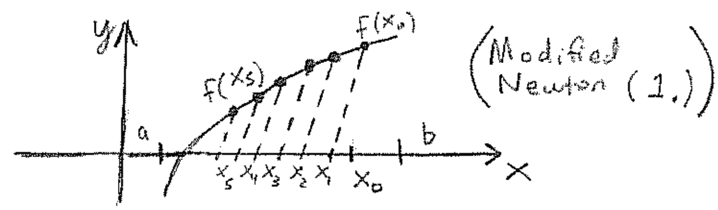
Notice that if $f(x_1) = 0$ then we found the zero and the method just gives $x_2 = x_1$. The idea then is to continue in this fashion and define the n-th guess iteratively by

Newton's Method:

$$\boxed{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}}$$

If for some particular n we actually find the exact value of the zero then the iteration just stays on that value. Otherwise, it can be shown that $\lim_{n\to\infty} x_n = x_*$ where $f(x_*) = 0$.





This is the simplest form of Newton's method but it is also perhaps the hardest to code. We'd have to calculate a new value for the derivative for each step. Edwards gives two modifications of the method and proves convergence for each.

Modified Newton Methods:

1. $\boxed{x_{n+1} = x_n - \frac{f(x_n)}{M}}$ where we know $0 < m < f'(x) < M$.

2. $\boxed{x_{n+1} = x_n - \frac{f(x_n)}{f'(a)}}$ where we know $f'(a) \neq 0$.

In case (1.) Edwards uses the concept of a **contraction mapping** to prove that the sequence converges and he even gives an estimate to bound the error of the guess (see Theorem 1.2 on pg. 164). Then he cautions against (2.) because it is possible to have Fig. 3.2 on pg. 162 occur, in other words if we guess badly to begin we might never find the root x_*. The remedy is fairly simple, you just look on smaller intervals. For (2.) he states the result concretely only in a local case (see Theorem 1.3 on pg. 165). I actually have only stated a particular case of his Theorem since I have made $b = 0$. The proof of the inverse function theorem builds from method (2.) but I'll give an example of (1.) because it's interesting and it should help make this whole discussion a little more tangible.

Example 9.1.1. Let $f(x) = \sin(x) + \frac{x}{2} - 1$. Find zero on $[-\pi/2, \pi/2]$. Notice $f'(x) = \cos(x) + \frac{1}{2}$ hence $f'(x) > 0$ for $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$. Moreover, $\frac{1}{2} \leq f'(x) \leq \frac{3}{2}$ for $x \in [-\pi/2, \pi/2]$.

Guess $x_o = 0$ and use $M = 3/2$.

$$x_1 = x_o - \tfrac{2}{3}f(x_o) = \tfrac{2}{3}.$$

$$x_2 = x_1 - \tfrac{2}{3}f(x_1) \cong 0.6989$$

$$x_3 = x_2 - \tfrac{2}{3}f(x_2) \cong 0.7037$$

$$x_4 = x_3 - \tfrac{2}{3}f(x_3) \cong 0.7044$$

$$x_5 = x_4 - \tfrac{2}{3}f(x_4) \cong 0.7045$$

last digit uncertain, $\boxed{x \cong 0.705}$

Remark: I had to zoom-in about 5 times on a TI-83 to verify this result

In case (2.) we can actually solve the equation $f(x) = y$ for a given value y close to b provided $f(a) = b$ and $f'(a) \neq 0$. The idea here is just to replace the function at $(x_o, f(x_o))$ with the line $L(x) = f(x_o) + f'(a)(x - x_o)$ and then we solve $L(x) = y$ to obtain $x = x_o - \frac{f(x_o)-y}{f'(a)}$. Note here we use the slope from the point (a, b) throughout the iteration, in particular we say $x_1 = x$ and start iterating as usual: $x_{n+1} = x_n - \frac{f(x_n)-y}{f'(a)}$ (see Theorem 1.3 on pg. 165 in Edwards for proof this converges)

Problem: given a function $f : \mathbb{R} \to \mathbb{R}$ which is continuously differentiable near a and $f'(a) \neq 0$, can we find a function g such that $f(g(y)) = y$ for y near the image $f(a)$?

Solution: Modified Newton's Method. we seek to solve $f(g(y)) = y$ for y in some neighborhood of a, simply define $g_o(y) = a$ and apply the method

$$\boxed{g_{n+1}(y) = g_n(y) - \frac{f(g_n(y)) - y}{f'(a)}}$$

Notice this can be done for each y near $f(a)$, in other words, we have a **sequence of functions** $\{g_n\}_{n=0}^{\infty}$. Moreover, if we take $n \to \infty$ this sequence **uniformly** converges to an exact solution g. This gives us an iterative way to construct local inverse functions for some given continuously differentiable function at a point a such that $f'(a) \neq 0$.

The idea of convergence of functions begs the question of what precisely is the "length" or "norm" of a function. Again, I postpone such discussion until the very end of the course. For now just accept that the idea of convergence of sequences of functions is well defined and intuitively it just means that the sequence matches the limiting function as $n \to \infty$. You encountered this idea in the discussion of Taylor series in calculus II, one can ask whether the sequence of Taylor polynomials for f does converge to f relative to some interval of convergence.

The calculations that follow here amaze me.

Example 9.1.2. (this is also in Edwards' Text)

$$f(x) = x^2 - 1, \quad a = 1, \quad b = 0 \Rightarrow f'(1) = 2(1) = 2.$$

$$g_0(y) = 1$$

$$g_1(y) = g_0(y) - \tfrac{1}{2}[f(g_0(y)) - y] = 1 - \frac{0-y}{2} = 1 + \tfrac{1}{2}y$$

$$g_2(y) = (1 + \tfrac{1}{2}y) - \tfrac{1}{2}[(1+y/2)^2 - y] = 1 + \tfrac{1}{2}y - \tfrac{1}{8}y^2$$

Thus the inverse of $f(x)$ near $a = 1$ is,

$$g(y) = 1 + \tfrac{1}{2}y - \tfrac{1}{8}y^2 + \cdots$$

Recall $(1+y)^k = 1 + ky + \frac{1}{2}k(k-1)y^2 + \cdots$ observe our $g(y)$ fits the binomial series with $k = 1/2$ thus

$$\boxed{g(y) = \sqrt{1+y} = f^{-1}(y)}$$ ← technically I should say $f|_{(-1,1)}^{-1}$

Example 9.1.3.

Usually this algorithm gives much weirder results. For example, $f(x) = \sin(x)$ take $(0,0)$ as base point note $f'(0) = \cos(0) = 1$.

$$g_0(y) = 0$$

$$g_1(y) = g_0(y) - [f(g_0(y)) - y] = 0 - \sin(0) + y = y.$$

$$g_2(y) = y - [\sin(y) - y] = 2y - \sin y$$

$$g_3(y) = 2y - \sin y - [\sin(2y - \sin y) - y]$$

Hence $\boxed{f^{-1}(y) \approx 3y - \sin(y) - \sin(2y - \sin y)}$

As a quick check try $f(g_3(y))$, use $\sin\theta \approx \theta$ for $\theta \approx 0$,

$$\begin{aligned} f(g_3(y)) &= \sin(3y - \sin y - \sin(2y - \sin y)) \\ &\cong \sin(3y - y - 2y + y) \\ &\cong \sin(y) \approx y. \end{aligned}$$

It would be interesting to implement this algorithm in Mathematica.

9.1.1 local solutions to level curves

Next, we try a similar technique to solve equations of the form $G(x,y) = 0$. You should recall that the solution set of $G(x,y) = 0$ is called a **level curve**. Usually we cannot make a global solution for y; in other words, there does not exist $f(x)$ such that $G(x, f(x)) = 0$ for all x in the solution set of G. For example, $G(x,y) = x^2 + y^2 - 1$ allows us to cast the unit circle as the solution set of the equation $G(x,y) = 0$. But, the unit circle is not the graph of a single function since it fails the vertical line test. Instead we need a pair of functions to cover the circle. Generally the situation can get quite complicated. Let's pause to notice there are two points where we cannot find a solution to $G(x,y) = 0$ on an open disk about the point: these points are $(-1,0)$ and $(1,0)$. We have trouble at the vertical tangents, note $G_y(x,y) = 2y$ has $G_y(-1,0) = G_y(1,0) = 0$ [1].

Idea: use the Newton's method approach to find solution, however, the approach here is slightly indirect. We'll use the mean value theorem to replace a function with its tangent line. Consider a fixed x_* near a then we have an function of y alone: $h(y) = G(x_*, y)$. Apply the mean value theorem to h for a y-value y_* such that point (x_*, y_*) has $G(x_*, y_*) = 0$,

$$G_y(x_*, b) = \frac{G(x_*, y_*) - G(x_*, b)}{y_* - b} = -\frac{G(x_*, b)}{y_* - b}$$

We can solve for y_* to obtain:

$$y_* = b - \frac{G(x_*, b)}{G_y(x_*, b)}$$

Define $f_o(x) = b$ and define $f_1(x)$ by

$$f_1(x) = f_o(x) - \frac{G(x, f_o(x))}{G_y(x, f_o(x))} \quad \text{and} \quad f_2(x) = f_1(x) - \frac{G(x, f_1(x))}{G_y(x, f_1(x))} \quad \text{and so forth...}$$

Foruntately, Edwards proves we can use an easier formula where the denomimator is replaced with $G_y(a,b)$ which is pretty close to the formula we have above provided the point considered is close to (a,b).

Theorem 9.1.4. *(Theorem 1.4 in Edwards's Text)*

Let $G : \mathbb{R}^2 \to \mathbb{R}$ be continuously differentiable and (a,b) a point such that $G(a,b) = 0$ and $G_y(a,b) \neq 0$ then we can find a function f on some closed interval J centered at a which covers the solution set of $G(x,y) = 0$ near all points close to (a,b). Moreover, this **local solution** is the limit of the sequence of functions inductively defined below:

$$f_o(x) = b \quad \text{and} \quad f_{n+1}(x) = f_n(x) - \frac{G(x, f_n(x))}{G_y(a,b)}$$

for all $n \in \mathbb{N}$. **We can calculate solutions iteratively!**

[1] yes, if we used closed disks then we could find a solution on a disk where $(-1,0)$ or $(1,0)$ was on the boundary, the point of the discussion is to motivate the implicit function theorem's langauge

Look at Example 2 on page 170 for a nice straight-forward application of Theorem 1.4. Perhaps you're not too excited by this example. Certainly, algebra solves the problem with ease anyway, we just have to take care with the algebraic steps. I intend for the next example to confound algebraic techniques and yet we can find an approximate solution:

Example 9.1.5. *Let $G(x,y) = exp(x^2+y^2)+x-e$. Notice that $G(0,1)=0$ and $G_y(0,1)=2$. Apply the algorithm:*

$$\begin{aligned} f_o(x) &= 1 \\ f_1(x) &= 1 - \tfrac{1}{2}G(x,1) = 1 - \tfrac{1}{2}(exp(x^2+1)+x-e) \\ f_2(x) &= f_1(x) - \tfrac{1}{2}[exp(x^2+[f_1(x)]^2+x-e] \end{aligned}$$

I'd go on but it just gets ugly. What is neat is that

$$y = f_1(x) = 1 - \tfrac{1}{2}(exp(x^2+1)+x-e)$$

gives an approximation of a local solution of $exp(x^2+y^2)+x-e=0$ for points near $(0,1)$.

Example 9.1.6. *Let $G(x,y) = x^2+y^2+y-1$ note $G_y = 2y+1$. Note that $G(1,0)=0$ and $G_y(1,0)=1$. Calculate the local solution by the algorthim:*

$$\begin{aligned} f_o(x) &= 0 \\ f_1(x) &= 0 - G(x,0) = 1-x^2 \\ f_2(x) &= 1-x^2 - G(x,1-x^2) = x^2-x^4 \\ f_3(x) &= x^2-x^4-G(x,x^2-x^4) = 1-x^2-x^4+2x^6-x^8 \end{aligned}$$

Now, these formulas are somewhat bizarre because we are writing an approximation centered at $x=1$ as polynomials centered at zero. It is probable that a nicer pattern emerges if we were to write all of these as polynomials in $(x-1)$. Notice that $f_n(1)=0$ for $n=0,1,2,3$.

Example 9.1.7. *Let $G(x,y) = x^2+y^2+y-2$ note $G_y = 2y+1$. Note that $G(0,1)=0$ and $G_y(0,1)=3$. Calculate the local solution by the algorthim:*

$$f_o(x) = 1$$

$$\begin{aligned} f_1(x) &= 1 - \tfrac{1}{3}G(x,1) \\ &= 1 - \tfrac{1}{3}x^2 \end{aligned}$$

$$\begin{aligned} f_2(x) &= 1 - \tfrac{1}{3}x^2 - G(x,1-\tfrac{1}{3}x^2) \\ &= 1 - \tfrac{1}{3}x^2 - [x^2+(1-\tfrac{1}{3}x^2)^2+(1-\tfrac{1}{3}x^2)-2] \\ &= 1 - \tfrac{1}{3}x^2 - \tfrac{1}{9}x^4 \end{aligned}$$

Note how the approximation unfolds order by order when the center matches the format in which we write the expansion.

If the center $a \neq 0$ then what can happen is that the terms of a particular order get spread across all orders in the Newton's Method approximation. I've found the expansions generated from the Newton's method are not easy to write in a nice form in general... of course, this shouldn't be that surprising, the method just gave us a way to solve problems that defy closed-form algebraic solution.

9.1.2 from level surfaces to graphs

In the preceding section we found that $G(x, y) = 0$ could be understood as a graph of a function of a single variable locally, in other words we found a 1-manifold. When we have an equation of n-variables it will likewise find $(n-1)$ free variables. This means that $G(x, y, z) = x^2+y^2+z^2-1 = 0$ gives us a level-surface (the sphere), or $G(t, x, y, z) = -t^2+x^2+y^2+z^2 = 0$ gives a level-volume (the light cone[2]). If we can solve the equation $G(x_1, x_2, \dots, x_n)$ for x_j then we say we have re-written the level surface as a graph. This is important because graphs are a special case of a parametrized manifold, the parametric formalism allows us to set-up integrals over higher-dimensional surfaces and so forth. These things will become clearer when we study integration of differential forms later in this course. I state Theorem 1.5 in Edwards here for completeness. The essential point is this, if $\nabla G(p) \neq 0$ then there exists j such that $\frac{\partial G}{\partial x_j}(p) \neq 0$ and we can solve for x_j by using basically the same the iterative process we just worked out in the $n = 2$ case in the preceding subsection.

Theorem 9.1.8. *(Theorem 1.5 in Edwards's Text)*

Let $G : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and $p = (a_1, a_2, \dots, a_n)$ a point such that $G(p) = 0$ and $G_j(p) \neq 0$ then we can find a function f on some closed interval J centered at a_j which covers the solution set of $G(x_1, x_2, \dots, x_n) = 0$ near all points close to p. Moreover, this **local solution** is the limit of the sequence of multivariate functions inductively defined below:

$$f_o(\vec{x}) = a_j \quad \text{and} \quad f_{n+1}(\vec{x}) = f_n(\vec{x}) - \frac{G(x_1, \dots, f(\vec{x}), \dots, x_n)}{G_{x_j}(p)}$$

for all $n \in \mathbb{N}$. If $f = \lim_{n\to\infty} f_n$ then $G(x_1, \dots, f(\vec{x}), \dots, x_n) = 0$ for points near p.

Something interesting happens when we apply this theorem to examples which allow explicit closed-form algebraic solution.

Example 9.1.9. *Consider* $G(x, y, z) = x+y+2z-4 = 0$. *Note that* $G_z = 2 \neq 0$ *and* $G(1, 1, 1) = 0$. *Apply the algorithm:*

$$\begin{aligned}
f_o(x, y) &= 1 \\
f_1(x, y) &= 1 - \tfrac{1}{2}G(x, y, 1) = 1 - \tfrac{1}{2}(x + y + 2 - 4) = -\tfrac{1}{2}(x + y - 4) \\
f_2(x, y) &= -\tfrac{1}{2}(x + y - 4) - \tfrac{1}{2}G(x, y, -\tfrac{1}{2}(x + y - 4)) = f_1(x, y)
\end{aligned}$$

You can clearly see that $f_n = f_1$ *for all* $n \geq 1$ *thus* $\lim_{n\to\infty} f_n = f_1$. *In other words, we found the exact solution is* $z = -\frac{1}{2}(x + y - 4)$.

[2] physically this represents the border of the spacetime which we can interact with in the future or the past, granting that special relativity actually describes nature without exception...

You might wonder if this just happened because the preceding example was linear, in fact, it has little to do with it. Here's another easy example,

Example 9.1.10. *Consider* $G(x,y,z) = x^2+y^2-z = 0$. *Note that* $G_z = -1 \neq 0$ *and* $G(0,0,0) = 0$. *Apply the algorithm:*

$$\begin{aligned} f_o(x,y) &= 0 \\ f_1(x,y) &= 0 + G(x,y,0) = x^2+y^2 \\ f_2(x,y) &= x^2+y^2+G(x,y,x^2+y^2) = x^2+y^2+[x^2+y^2-(x^2+y^2)] = f_1(x,y) \end{aligned}$$

You can clearly see that $f_n = f_1$ *for all* $n \geq 1$ *thus* $\lim_{n\to\infty} f_n = f_1$. *In other words, we found the exact solution is* $z = x^2+y^2$.

Part of the reason both of the preceding examples were easy is that the solutions were not just local solutions, in fact they were global. When the solution is the level surface equation breaks up into cases it will be more complicated.

Example 9.1.11. *Suppose* $G(x,y,z) = \sin(x+y-z) = 0$ *then solutions must satisfy* $x+y-z = n\pi$ *for* $n \in \mathbb{Z}$. *In other words, the algorithm ought to find* $z = x+y-n\pi$ *where the choice of* n *depends on the locality we seek a solution. This level-set is actually a whole family of disconnected paralell planes. Let's see how the algorithm deals with this, feed it* $(0,0,2\pi)$ *as the starting point (this ought to select the* $n = -2$ *surface. Apply the algorithm to* $G(x,y,z) = \sin(x+y-z)$ *where clearly* $G(0,0,2\pi) = 0$ *and* $G_z = -\cos(-2\pi) = -1$ *hence:*

$$\begin{aligned} f_o(x,y) &= 2\pi \\ f_1(x,y) &= 2\pi + G(x,y,2\pi) = 2\pi + \sin(x+y+2\pi) = 2\pi+\sin(x+y) \\ f_2(x,y) &= 2\pi + \sin(x+y) + \sin(x+y+\sin(x+y)) \\ f_3(x,y) &= 2\pi + \sin(x+y) + \sin(x+y+\sin(x+y)) \\ &\quad + \sin(x+y+\sin(x+y)+\sin(x+y+\sin(x+y))) \end{aligned}$$

I deem these formulas weird. Perhaps I can gain some insight by expanding f_1,

$$f_1(x,y) = 2\pi + x + y - \tfrac{1}{3!}(x+y)^3 + \cdots$$

I'm a little scared to look at f_2. *There must be some sort of telescoping that happens in order for us to obtain the real solution of* $z = x+y+2\pi$.

It's not at all obvious to me how the formula above telescopes in the limit that $n \to \infty$. However, unless I'm missing something or making a silly mistake, it seems clear that G is continuously differentiable at $(0,0,2\pi)$ and $G_z(0,0,2\pi) \neq 0$. Therefore, Theorem 1.5 applies and the sequence of function f_n should uniformly converge to the solution we know exists through direct argument in this example. Anyway, my point in this section is not to make a blanket endorsement that you solve all equations by the algorthim. I am merely trying to illustrate how it works.

9.2 inverse and implicit mapping theorems

In the preceding section we began by motivating the inverse function theorem for functions of one variable. In short, if the derivative is nonzero at a point then the function is 1-1 when restricted to a neighborhood of the point. Newton's method, plus a bunch of careful analysis about contraction mappings which we skipped this semester, then gave an algorithm to calculate the local inverse for a function of one variable. After that we essentially applied the local inverse idea to the problem of solving a level curve $G(x,y) = 0$ locally for an explicit solution of y. The result that such a solution is possible near points where $G_y \neq 0$ is known as the **implicit function theorem.** We then concluded by observing that almost the same mathematics allowed us to find an explicit solution of $G(x_1, \dots, x_{n+1}) = 0$ for one of the variables provided the partial derivative in that direction was nonzero. This result is also called the **implicit function theorem.** We used these theorems implicitly when I pulled parametrizations from my imagination, typically it is the implicit function theorem that justifies such a step. Moreover, to insist $\nabla g(p) \neq 0$ means that there exists at least one partial derivative nonzero so the implicit function theorem applies. All of that said, this section is basically the same story again. Difference is we have to deal with a little extra notation and linear algebra since a mapping is actually an ensemble of functions dealt with at once.

9.2.1 inverse mapping theorem

Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ has an inverse $f^{-1} = g$ then we have $f \circ g = Id$ so the chain rule yields $df \circ dg = d(Id) = Id$ since the identity is a linear map and hence it is its own best linear approximation. Note that we find that $f'g' = I_n$ thus $(f')^{-1} = g'$ or in other notation $[f']^{-1} = [f^{-1}]'$. With this in mind we wish to find a formula to calculate the inverse function. The definition seems like a good place to start:

$$\begin{aligned} f(g(y)) = y \quad &\Rightarrow \quad g(y) = f^{-1}(y) \\ &\Rightarrow \quad g(y) \approx g(f(a)) + g'(a)[y - f(a)] \\ &\Rightarrow \quad g(y) \approx a + [f'(a)]^{-1}[y - f(a)] \\ &\Rightarrow \quad g_1(y) = g_o(y) + [f'(a)]^{-1}[y - f(g_o(y))] \quad \text{where} \quad g_o(y) = a \\ &\Rightarrow \quad g_{n+1}(y) = g_n(y) + [f'(a)]^{-1}[y - f(g_n(y))] \quad \text{where} \quad g_o(y) = a \end{aligned}$$

Theorem 9.2.1. *(Theorem 3.3 in Edwards's Text see pg 185)*

Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable in an open set W containing a and the derivative matrix $f'(a)$ is invertible. Then f is locally invertible at a. This means that there exists an open set $U \subseteq W$ containing a and V a open set containing $b = f(a)$ and a one-one, continuously differentiable mapping $g : V \to W$ such that $g(f(x)) = x$ for all $x \in U$ and $f(g(y)) = y$ for all $y \in V$. Moreover, the local inverse g can be obtained as the limit of the sequence of successive approximations defined by

$$g_o(y) = a \quad \text{and} \quad g_{n+1}(y) = g_n(y) - [f'(a)]^{-1}[f(g_n(y)) - y]$$

for all $y \in V$.

Notice this theorem gives us a way to test coordinate mappings for invertibility, we can simply calculate the derivative matrix then calculate its determinant to check to see it is nonzero to insure invertibility and hence the local invertibility of the coordinate map. There still remains the danger that the mapping doubles back to the same value further out so if we insist on a strict one-one correspondance then more analysis is needed to make sure a given transformation is indeed a coordinate system. (see Ex 1 on pg. 183 for a function which is everywhere locally invertible and yet not an injective mapping)

Example 9.2.2.

$$f(r,\theta) = (r\cos\theta,\ r\sin\theta) \longrightarrow f'(r,\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{bmatrix}$$

$$\det\big(f'(r,\theta)\big) = r\cos^2\theta + r\sin^2\theta = r.$$

It follows f is locally invertible everywhere except the origin. We know from previous experience $\cos(\theta + 2\pi) = \cos\theta$ spoils the existence of a global inverse on $\mathbb{R}^2 - \{(0,0)\}$.

Example 9.2.3.

$f(\vec{v}) = \vec{v} \times \vec{a}$ where $\vec{a} \neq 0$ is a fixed vector in $\mathbb{R}^3$ thus $f : \mathbb{R}^3 \to \mathbb{R}^3$. Let $\vec{a} = \langle a_x, a_y, a_z \rangle$

$$f(x,y,z) = \langle x,y,z \rangle \times \langle a_x, a_y, a_z \rangle$$

$$= \langle \underbrace{y a_z - z a_y}_{f_1},\ \underbrace{z a_x - x a_z}_{f_2},\ \underbrace{x a_y - y a_x}_{f_3} \rangle$$

$$f'(x,y,z) = \begin{bmatrix} 0 & a_z & -a_y \\ -a_z & 0 & a_x \\ a_y & -a_x & 0 \end{bmatrix}$$

$$\det\left[f'(x,y,z)\right] = a_z a_x a_y - a_y a_x a_z = 0.$$

Therefore, f is nowhere invertible. For example, $\vec{a} = \langle 0,0,1 \rangle$

$$f(\vec{v}) = \vec{v} \times \langle 0,0,1 \rangle = \langle v_y, -v_x, 0 \rangle$$

← for $\vec{a} = \langle 0,0,1 \rangle$ this always happens.

9.2.2 implicit mapping theorem

Let me begin by stating the problem we wish to consider:

Given continuously differentiable functions $G_1, G_2, \dots, G_n$

$$\begin{aligned} G_1(x_1, \dots, x_m, y_1, \dots, y_n) &= 0 \\ G_2(x_1, \dots, x_m, y_1, \dots, y_n) &= 0 \\ &\vdots \\ G_n(x_1, \dots, x_m, y_1, \dots, y_n) &= 0 \end{aligned}$$

Locally solve $y_1, \dots, y_n$ as functions of $x_1, \dots x_m$. That is, find a mapping $h : \mathbb{R}^m \to \mathbb{R}^n$ such that $G(x, y) = 0$ iff $y = h(x)$ near some point $(a, b) \in \mathbb{R}^{m+n}$ such that $G(a, b) = 0$. In this section we use the notation $x = (x_1, x_2, \dots x_m)$ and $y = (y_1, y_2, \dots, y_n)$.

It is convenient to define partial derivatives with respect to a whole vector of variables,

$$\frac{\partial G}{\partial x} = \begin{bmatrix} \frac{\partial G_1}{\partial x_1} & \cdots & \frac{\partial G_1}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial G_n}{\partial x_1} & \cdots & \frac{\partial G_n}{\partial x_m} \end{bmatrix} \qquad \frac{\partial G}{\partial y} = \begin{bmatrix} \frac{\partial G_1}{\partial y_1} & \cdots & \frac{\partial G_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial G_n}{\partial y_1} & \cdots & \frac{\partial G_n}{\partial y_n} \end{bmatrix}$$

Consider $h : \mathbb{R}^m \to \mathbb{R}^n$ such that $G(x, y) = 0$ iff $y = h(x)$ near some point $(a, b) \in \mathbb{R}^{m+n}$ such that $G(a, b) = 0$. In other words, suppose $G(x, h(x)) = 0$. The chain rule reads:

$$0 = \frac{\partial G}{\partial x} + \frac{\partial G}{\partial y} h'(x)$$

Or, provided the matrix $\frac{\partial G}{\partial y}$ is invertible we can calculate,

$$h'(x) = -\left[\frac{\partial G}{\partial y}\right]^{-1} \frac{\partial G}{\partial x}$$

Theorem 9.2.4. *(Theorem 3.4 in Edwards's Text see pg 190)*

Let $G : \mathbb{R}^{n+m} \to \mathbb{R}^n$ be continuously differentiable in a open ball about the point (a, b) where $G(a, b) = 0$. If the matrix $\frac{\partial G}{\partial y}(a, b)$ is invertible then there exists an open ball U containing a in $\mathbb{R}^m$ and an open ball W containing (a, b) in $\mathbb{R}^{n+m}$ and a continuously differentiable mapping $h : U \to \mathbb{R}^n$ such that $G(x, y) = 0$ iff $y = h(x)$ for all $(x, y) \in W$. Moreover, the mapping h is the limit of the sequence of successive approximations defined inductively below

$$h_o(x) = b, \quad h_{n+1} = h_n(x) - [\tfrac{\partial G}{\partial y}(a, b)]^{-1} G(x, h_n(x))$$

for all $x \in U$.

I have given barely enough details to understand the notation here. If you read pages 188-194 of Edwards you can have a much deeper understanding. I will not attempt to recreate his masterpiece here. One important notation I should mention is the so-called Jacobian of G with respect to y. It is the determinant of the partial derivative matrix $\frac{\partial G}{\partial y}$ which is denoted $det\frac{\partial G}{\partial y} = \frac{\partial(G_1,G_2,\dots,G_n)}{\partial(y_1,y_2,\dots,y_n)}$. This gives us an easy criteria to check on the invertibility of $\frac{\partial G}{\partial y}$. Note that if this Jacobian is nonzero then we may judge the level set $G(x,y)=0$ is an n-dimensional space since it is in one-one correspondence of some open ball in $\mathbb{R}^n$.

Remark 9.2.5.

You may recall the strange comments in red from my section 6.2. I discussed the rank of various derivative matrices. In this section we put the free variables (x) at the start of the list and the dependent variables (y) at the end, however, this is just a notational choice. In practice if we can select any set of n-variables for $G(z_1,z_2,\dots,z_{m+n})=0$ such that $det[G_{i_1}|G_{i_2}|\cdots|G_{i_n}] \neq 0$ then we can solve for $z_{i_1},\dots,z_{i_n}$ in terms of the remaining variables. Thus, in retrospect, showing full rank of the derivative matrix could justifies the local invertibility of certain mappings.

Example 9.2.6. (using Ex. 6.2.10 from my Chapter 6)

$F:\mathbb{R}^3 \to \mathbb{R}^2$ where $F(x,y,z) = (x^2+z^2, yz)$

$$\frac{\partial(F_1,F_2)}{\partial(x,y)} = det\begin{bmatrix} \partial F_1/\partial x & \partial F_1/\partial y \\ \partial F_2/\partial x & \partial F_2/\partial y \end{bmatrix} = det\begin{bmatrix} 2x & 0 \\ 0 & z \end{bmatrix} = 2xz$$

This means we can solve $F(x,y,z) = (0,0)$ in terms of z at points where $2xz \neq 0$.
We can find $f_1(z)$ and $f_2(z)$ such that
$F(f_1(z), f_2(z), z) = 0$. (I see $f_1(z) = -z^2$, $f_2(z) = 0$).

$F = 0$ gives curve $\vec{r}(z) = \langle -z^2, 0, z\rangle$

Example 9.2.7.

Same function, different choice of free parameter,

$$\frac{\partial(F_1,F_2)}{\partial(y,z)} = det\begin{bmatrix} F_{1y} & F_{1z} \\ F_{2y} & F_{2z} \end{bmatrix} = det\begin{bmatrix} 0 & 2z \\ z & y \end{bmatrix} = -2z^2$$

We can solve $F=0$ where $z \neq 0$ in terms of the remaining variable x; find $y = f_1(x)$ and $z = f_2(x)$ such that $F(x, f_1(x), f_2(x)) = (0,0)$. A little thinking reveals $f_1(x) = 0$ and $f_2(x) = -x^2$ hence we've found $F=0$ has sol$^{\underline{n}}$ $\vec{r}(x) = \langle x, 0, -x^2\rangle$ (parametrization of sol$^{\underline{n}}$.)

9.3 implicit differentiation

Enough theory, let's calculate. In this section I apply previous theoretical constructions to specific problems. I also introduce standard notation for "constrained" partial differentiation which is also sometimes called "partial differentiation with a side condition".

Example 9.3.1. Suppose $\overbrace{xyz + 2x^2z + 3xz^2 - 1}^{G} = 0$. Calculate $\left(\frac{\partial z}{\partial x}\right)_y$ or $\left(\frac{\partial z}{\partial y}\right)_x$. Here the notation $\left(\frac{\partial z}{\partial x}\right)_y$ announces y is independent from x. We have $G(x,y,z) = 0$ thus $dG = \frac{\partial G}{\partial x}dx + \frac{\partial G}{\partial y}dy + \frac{\partial G}{\partial z}dz$ and $\frac{\partial G}{\partial z} = xy + 2x^2 + 6xz \neq 0$ thus the implicit function Th^{m} applies and we can find $z = h(x,y)$ for some function h at least locally where $G_z \neq 0$

$$0 = G_x\,dx + G_y\,dy + G_z\,dz \quad , \text{ if } G_z \neq 0 \text{ then}$$

$$dz = \frac{-G_x}{G_z}dx - \frac{G_y}{G_z}dy \implies \left(\frac{\partial z}{\partial x}\right)_y = \frac{-G_x}{G_z} = \frac{-yz - 4xz - 3z^2}{xy + 2x^2 + 6xz}.$$

Example 9.3.2.

Likewise, $dz = \left(\frac{\partial z}{\partial x}\right)_y dx + \left(\frac{\partial z}{\partial y}\right)_x dy$ hence,

$$\left(\frac{\partial z}{\partial y}\right)_x = \frac{-G_y}{G_z} = \frac{-xz}{xy + 2x^2 + 6xz} = \frac{-z}{y + 2x + 6z}.$$

The idea is this: the implicit function Th^{m} gives us the existence of a sol^{n} to $G(x_1, \dots, x_j, \dots, x_n) = 0$ provided $G_j \neq 0$ then partial derivatives w.r.t. the variables $x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_n$ can be obtained by solving $dG = G_1 dx_1 + \cdots + G_j dx_j + \cdots + G_n dx^n$ for dx_j.

Example 9.3.3. Same G, think of x, z as independent & y dependent,

$$dy = \frac{-G_x}{G_y}dx - \frac{G_z}{G_y}dz$$

$$\left(\frac{\partial y}{\partial z}\right)_x = \frac{-G_z}{G_y} = \frac{-xy - 2x^2 - 6xz}{xz}.$$

Example 9.3.4. Let $\begin{cases} 2x + y - 3z - 2u = 0 \\ x + 2y + z + u = 0 \end{cases}$ find $\left(\frac{\partial x}{\partial y}\right)_z$.

Note we have 2 eq^{ns} and 4 unknowns. The implicit mapping Th^{m} may provide an implicit sol^{n} locally. Notice y, z are free if we are to calculate $\left(\frac{\partial x}{\partial y}\right)_z$ hence we should suspect $(x, u) = h(y, z)$. $(h : \mathbb{R}^2 \longrightarrow \mathbb{R}^2)$

$$\text{(I)} \quad 2dx + dy - 3dz - 2du = 0$$
$$\text{(II)} \quad dx + 2dy + dz + du = 0$$

Add (I) and 2(II) to obtain (trying to eliminate du)

$$4dx + 5dy - dz = 0$$
$$\therefore \ dx = \frac{-5}{4}dy - \frac{1}{4}dz \implies \boxed{\left(\frac{\partial x}{\partial y}\right)_z = \frac{-5}{4}}$$

Example 9.3.5.

Generally suppose
$$G_1(x_1, \dots, x_m, y_1, \dots, y_n) = 0$$
$$G_2(\vec{x}, \vec{y}) = 0$$
$$\vdots$$
$$G_n(\vec{x}, \vec{y}) = 0$$

We have n-eq^{n}'s and $(m+n)$-unknowns. We can solve for $\vec{y}$ in terms of $\vec{x}$ provided $\frac{\partial G}{\partial \vec{y}}$ is invertible. This is equivalent to saying we can solve for $d\vec{y}$. the eq^{n} $G_{\vec{x}}\,d\vec{x} + G_{\vec{y}}\,d\vec{y}$ we need $G_{\vec{y}} = \frac{\partial G}{\partial \vec{y}}$ to have $\det\left(\frac{\partial G}{\partial \vec{y}}\right) = \frac{\partial(G_1, \dots, G_n)}{\partial(y_1, \dots, y_n)} \neq 0$.

Again we wish to solve for the dy's,

$$dG_1 = \frac{\partial G_1}{\partial x_1}dx_1 + \cdots + \frac{\partial G_1}{\partial x_m}dx_m + \frac{\partial G_1}{\partial y_1}dy_1 + \cdots + \frac{\partial G_1}{\partial y_n}dy_n = 0$$
$$\vdots$$
$$dG_n = \frac{\partial G_n}{\partial x_1}dx_1 + \cdots + \frac{\partial G_n}{\partial x_m}dx_m + \frac{\partial G_n}{\partial y_1}dy_1 + \cdots + \frac{\partial G_n}{\partial y_n}dy_n = 0$$

$$\begin{bmatrix} \partial G_1/\partial y_1 & \partial G_1/\partial y_2 & \cdots & \partial G_1/\partial y_n \\ \partial G_2/\partial y_1 & - & - & \partial G_2/\partial y_n \\ \vdots & & & \vdots \\ \partial G_n/\partial y_1 & - & - & \partial G_n/\partial y_n \end{bmatrix} \begin{bmatrix} dy_1 \\ dy_2 \\ \vdots \\ dy_n \end{bmatrix} = \begin{bmatrix} -\frac{\partial G_1}{\partial x_1}dx_1 - \cdots - \frac{\partial G_1}{\partial x_m}dx_m \\ \vdots \\ -\frac{\partial G_n}{\partial x_1}dx_1 - \cdots - \frac{\partial G_n}{\partial x_m}dx_m \end{bmatrix}$$

$$\left[\frac{\partial G}{\partial y_1} \middle| \frac{\partial G}{\partial y_2} \middle| \cdots \middle| \frac{\partial G}{\partial y_n}\right]\left[d\vec{y}\right] = \left[-\frac{\partial G}{\partial \vec{x}}d\vec{x}\right]$$

Example 9.3.6. I'll solve for dy_1

$$dy_1 = \frac{\det\left[-\frac{\partial G}{\partial \vec{x}} d\vec{x} \,\middle|\, \frac{\partial G}{\partial y_2} \,\middle|\, \cdots \,\middle|\, \frac{\partial G}{\partial y_n} \right]}{\det\left[\frac{\partial G}{\partial \vec{y}} \right]}$$

$$dy_2 = \frac{\det\left[\frac{\partial G}{\partial y_1} \,\middle|\, -\frac{\partial G}{\partial \vec{x}} d\vec{x} \,\middle|\, \frac{\partial G}{\partial y_3} \,\middle|\, \cdots \,\middle|\, \frac{\partial G}{\partial y_n} \right]}{\det\left[\frac{\partial G}{\partial \vec{y}} \right]}$$

etc... (this notation is unfolded on the previous page, see Ex. 9.3.5.)

Example 9.3.7.