LECTURE 10

Date of Lecture: February 17, 2020

As always, $\mathbf{K} \in \{\mathbf{R},\, \mathbf{C}\}$.

The symbol ☡ is for flagging a cautionary comment or a tricky argument. It occurs in the margins and is Knuth's version of Bourbaki's "dangerous bend symbol".

An n-tuple $(x_1, \dots, x_n)$ of symbols (x_i not necessarily real or complex numbers) will also be written as a column vector when convenient. Thus

$$(x_1, \dots, x_n) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

A map $\boldsymbol{f}$ from a set S to a product set $T_1 \times \cdots \times T_n$ will often be written as an n-tuple $\boldsymbol{f} = (f_1, \dots, f_n)$, with f_i a map from S to T_i, and hence, by the above convention, as a column vector

$$\boldsymbol{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}.$$

(See Remark 2.2.2 of Lecture 5.)

The default norm on Euclidean spaces of the form $\mathbf{R}^n$ is the Euclidean norm $\|\ \|_2$ and we will simply denote it as $\|\ \|$.

☡ Note that $(x_1, \dots, x_n) \neq [x_1 \dots x_n]$. Each side is the transpose of the other.

1. The Inverse Function Theorem

1.1. Derivatives of linear maps.

Let $T\colon \mathbf{R}^n \to \mathbf{R}^m$ be an $\mathbf{R}$-linear map. Then for $\boldsymbol{a}$ and $\boldsymbol{h}$ in $\mathbf{R}^n$ we have $T(\boldsymbol{a}+\boldsymbol{h}) - T(\boldsymbol{a}) - T(\boldsymbol{h}) = \mathbf{0}$ and so

$$\lim_{\boldsymbol{h}\to\mathbf{0}} \frac{\|T(\boldsymbol{a}+\boldsymbol{h}) - T(\boldsymbol{a}) - T\boldsymbol{h}\|}{\|\boldsymbol{h}\|} = 0. \tag{1.1.1}$$

We therefore get the following.

Proposition 1.1.2. *Let $T\colon \mathbf{R}^n \to \mathbf{R}^m$ be a linear map. Then $(DT)(\boldsymbol{a}) = T$ for all $\boldsymbol{a} \in \mathbf{R}^n$, $D^kT \equiv 0$ for all $k > 1$ and hence T is in $\mathscr{C}^\infty(\mathbf{R}^n) := \cap_{k\geq 0}\mathscr{C}^k(\mathbf{R}^n)$.*

Proof. From (1.1.1) it is clear that $(DT)(\boldsymbol{a}) = T$ for all $\boldsymbol{a} \in \mathbf{R}^n$. Thus $DT\colon \mathbf{R}^n \to L(\mathbf{R}^n,\, \mathbf{R}^m)$ is a constant. It follows that all its higher derivatives as zero. ☐

Corollary 1.1.3. *If $m = n$ and T is invertible, then T and T^{-1} are both in $\mathscr{C}^\infty(\mathbf{R}^n)$.*

1.2. **The formal statement of the theorem.** This is what we essentially proved in the last class:

Theorem 1.2.1. *Let* U *be an open subset in* $\mathbf{R}^n$, $\boldsymbol{f}\colon U \to \mathbf{R}^n$ *a* $\mathscr{C}^1$ *map,* $\boldsymbol{a} \in U$ *a point such that* $\boldsymbol{f}'(\boldsymbol{a})$ *is invertible, and let* $\boldsymbol{b} = \boldsymbol{f}(\boldsymbol{a})$*. Then there exists an open neighbourhood* V *of* $\boldsymbol{a}$ *in* U*, an open neighbourhood* W *of* $\boldsymbol{b}$ *in* $\mathbf{R}^n$*, such that* $W = \boldsymbol{f}(V)$, $\boldsymbol{f}|_V : V \to W$ *is one-to-one, and the inverse function* $\boldsymbol{\varphi}^{-1}\colon W \to V$ *of* $\boldsymbol{f}|_V$ *is also* $\mathscr{C}^1$.

Proof. Let $T = \boldsymbol{f}'(\boldsymbol{a})$. Let $\boldsymbol{g}$ the composite $U \xrightarrow{\boldsymbol{f}} \mathbf{R}^n \xrightarrow{T^{-1}} \mathbf{R}^n$. Then $\boldsymbol{g}'(\boldsymbol{a}) = I$ by the chain rule and Proposition 1.1.2, and clearly if the theorem is true for $\boldsymbol{g}$, it is true for $\boldsymbol{f}$. Thus, without loss of generality, we may assume $\boldsymbol{f}'(\boldsymbol{a}) = I$. Next replacing $\boldsymbol{f}$ with the function $\boldsymbol{\gamma}\colon U - \boldsymbol{a} \to \mathbf{R}^n$ given by $\boldsymbol{\gamma}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}+\boldsymbol{a}) - \boldsymbol{b}$ if necessary, we may assume without loss of generality that $\boldsymbol{a} = \boldsymbol{b} = \mathbf{0}$. In the last lecture, we proved the theorem in this special case. □

Corollary 1.2.2. *If* V *and* W *are as in the conclusion of the theorem, then* $\boldsymbol{f}'$ *and* $\boldsymbol{\varphi}'$ *are invertible at all points of* V *and* W *repsectively and* $\boldsymbol{\varphi}'(\boldsymbol{y}) = \boldsymbol{f}'(\boldsymbol{\varphi}(\boldsymbol{y}))^{-1}$ *for* $\boldsymbol{y} \in W$.

Proof. This occurs in the proof of the theorem. But it can be deduced without going through the proof by applying the chain rule for differentiation to $\boldsymbol{f}\circ\boldsymbol{\varphi}$ or to $\boldsymbol{\varphi}\circ\boldsymbol{f}$. □

1.3. **The Implicit Function Theorem.** Here is an important corollary of the Inverse Function Theorem—so important that it gets the status of a theorem rather than a corollary (in fact one can go the other way and prove the Inverse Function Theorem from the Implicit Function Theorem).

Theorem 1.3.1. *Let* d *and* m *be non-negative integers,* $n = d + m$*, and write all points of* $\mathbf{R}^n = \mathbf{R}^d \times \mathbf{R}^m$ *in the form* $(\boldsymbol{x}, \boldsymbol{y})$ *with* $\boldsymbol{x} \in \mathbf{R}^d$ *and* $\boldsymbol{y} \in \mathbf{R}^m$*. Let* U *be an open subset of* $\mathbf{R}^n$, $\boldsymbol{p} = (\boldsymbol{a}, \boldsymbol{b})$ *a point in* U*, and* $\boldsymbol{\varphi}\colon U \to \mathbf{R}^m$ *a* $\mathscr{C}^1$ *map such that the* $m \times m$ *matrix*

$$\frac{\partial \boldsymbol{\varphi}(\boldsymbol{p})}{\partial \boldsymbol{y}} := \begin{bmatrix} D_{d+1}\boldsymbol{\varphi}(\boldsymbol{p}) & D_{d+2}\boldsymbol{\varphi}(\boldsymbol{p}) & \dots & D_n\boldsymbol{\varphi}(\boldsymbol{p}) \end{bmatrix}$$

is invertible. Let $\boldsymbol{c} = \boldsymbol{\varphi}(\boldsymbol{p})$*. Then there exists an open neighbourhood* W *of* $\boldsymbol{a}$ *in* $\mathbf{R}^d$*, and a unique map* $\mathscr{C}^1$ *map* $\boldsymbol{f}\colon W \to \mathbf{R}^m$ *with the following properties*

- (a) $\boldsymbol{f}(\boldsymbol{a}) = \boldsymbol{b}$,
- (b) $(\boldsymbol{x}, \boldsymbol{f}(\boldsymbol{x})) \in U$ *for all* $\boldsymbol{x} \in W$,
- (c) $\boldsymbol{f}$ *is an implicit solution of the equation* $\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{c}$, *i.e.*

$$\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{f}(\boldsymbol{x})) = \boldsymbol{c} \qquad (\boldsymbol{x} \in W).$$

Proof. Without loss of generality, we may assume that $\boldsymbol{p} = \mathbf{0}$ and $\boldsymbol{c} = \mathbf{0}$. In other words, $\boldsymbol{a} = \mathbf{0}$, $\boldsymbol{b} = \mathbf{0}$, and $\boldsymbol{c} = \mathbf{0}$. We can do this by replacing, if necessary, $\boldsymbol{\varphi}$ by the map $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{\varphi}(\boldsymbol{x}+\boldsymbol{a}, \boldsymbol{y}+\boldsymbol{b}) - \boldsymbol{c}$ on the open set $U - \boldsymbol{p}$. The map $\Psi\colon U \to \mathbf{R}^n$ given by $\Psi(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{y}))$ has as its Jacobian matrix the block matrix

$$J\Psi = \begin{bmatrix} I & 0 \\ \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{x}} & \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{y}} \end{bmatrix}.$$

where I is the identity $d \times d$ matrix, and $\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{x}}$, $\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{y}}$ are the obvious $m \times d$ and $m \times m$ matrices respectively. It is clear from our hypotheses that $J\Psi(\mathbf{0})$ is invertible and hence the inverse function theorem applies to Ψ. Without loss of generality (by shrinking U around $\mathbf{0}$ if necessary) we may therefore assume that $V = \Psi(U)$ is open and $\Psi\colon U \to V$ is bijective and $\Psi^{-1}\colon V \to U$ is $\mathscr{C}^1$. Note that $\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{y}}$ is invertible on U, since the whole matrix $J\Psi$ is.

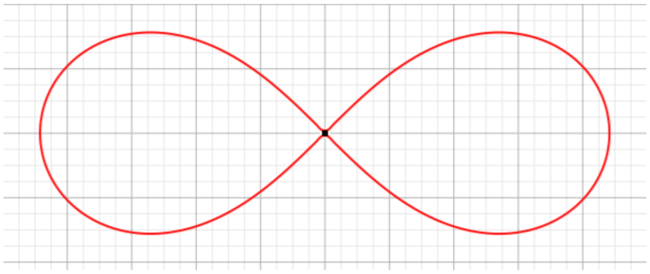
We have seen from our solution to Problem (1) in Quiz 3 that an implicit solution to $\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0}$ exists as a continuous solution in a neighbourhood W of $\mathbf{0} \in \mathbf{R}^d$. Here are the main points of the solution. Let $W \times W'$ be an open subset of U containing $\mathbf{0}$ where $W \subset \mathbf{R}^d$ and $W' \subset \mathbf{R}^m$. Such a rectangle obviously exists. We may assume without loss of generality that W is connected, i.e. W cannot be written as the disjoint union of two non-empty sets. Let $\boldsymbol{G} = \Psi^{-1}|_{W \times W'}$. Then $\boldsymbol{G} = (I, \boldsymbol{F})$ where $\boldsymbol{F}\colon W \times W' \to \mathbf{R}^m$ is a continuous map. In fact, it is $\mathscr{C}^1$ since $\boldsymbol{G}$ is. In greater detail, $J\boldsymbol{F}$ is a submatrix of $J\boldsymbol{G}$ and all entries in $J\boldsymbol{G}$ are continuous functions since $\boldsymbol{G}$ is $\mathscr{C}^1$. We are using Theorem 2.2.1 of Lecture 8 to draw the conclusion. And then another application of the same theorem (in the converse direction) shows that $\boldsymbol{F}$ is $\mathscr{C}^1$. Moreover, $\boldsymbol{f}(\boldsymbol{x}) = \pi'(\boldsymbol{G}(\boldsymbol{x}, \mathbf{0}))$, for $\boldsymbol{x} \in W$, where $\pi'\colon \mathbf{R}^d \times \mathbf{R}^m \to \mathbf{R}^m$ is the projection map $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{y}$. One more application of Theorem 2.2.1 of Lecture 8 shows that $\boldsymbol{f}$ is $\mathscr{C}^1$. Finally suppose $\boldsymbol{g}\colon W \to \mathbf{R}^n$ is another implicit solution of $\boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0}$, with $(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}))$ with $\boldsymbol{g}(\mathbf{0}) = \mathbf{0}$. Let

$$W_0 = \{\boldsymbol{x} \in W \mid (\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x})) \in W \times W'\}.$$

Then $\mathbf{0} \in W_0$ and hence W_0 is a non-empty and open. Then we have the chain of equalities: $(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x})) = \boldsymbol{G}(\Psi(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}))) = \boldsymbol{G}(\boldsymbol{x}, \boldsymbol{\varphi}(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}))) = \boldsymbol{G}(\boldsymbol{x}, \mathbf{0}) = (\boldsymbol{x}, \boldsymbol{f}(\boldsymbol{x}))$ for $\boldsymbol{x} \in W_0$. Thus $\boldsymbol{g}$ and $\boldsymbol{f}$ agree locally around $\mathbf{0}$. If we take any point $\boldsymbol{e} \in W$ on which $\boldsymbol{f}$ and $\boldsymbol{g}$ agree, then translating the orgin to $\boldsymbol{e}$ if necessary, we see that $\boldsymbol{f}$ and $\boldsymbol{g}$ agree in a neighbourhood of $\boldsymbol{e}$. Indeed, we have observed that $\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{y}}$ is invertible on W, and hence in particular at $\boldsymbol{e}$, allowing us repeat the argument we gave when we assumed $\boldsymbol{f}(\mathbf{0}) = \boldsymbol{g}(\mathbf{0})$. What we have shown is that the set $S = \{\boldsymbol{x} \in W \mid \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})\}$ is non-empty and open in W. It is clearly closed too. Since W is connected, this means $W \smallsetminus S = \varnothing$. Thus $S = W$ and $\boldsymbol{g} = \boldsymbol{f}$. ☐

Examples 1.3.2. You have already come across the implicit function theorem in high school. Whenever you differentiated implicitly, you used the implicit function theorem, and for most of you, without knowing it.

1. Consider the equation $2(x^2 + y^2)^2 - 25(x^2 - y^2) = 0$. Here is the graph.



This is of the form $\varphi(x, y) = 0$ where $\varphi\colon \mathbf{R}^2 \to \mathbf{R}$ is the function $\varphi(x, y) = 2(x^2 + y^2)^2 - 25(x^2 - y^2)$. In what sense can we find an implicit function f such that $\varphi(x, f(x)) \equiv 0$? What should be the domain of f? To get at the answer, we have to make sure the hypotheses of the implicit function theorem are satisfied. To that end first find the intervals where $\frac{\partial \varphi}{\partial y} \neq 0$. It is easier to find the locus where $\frac{\partial\varphi}{\partial y} = 0$. Thus we set

$$8(x^2 + y^2)y + 50y = 0.$$

This yields

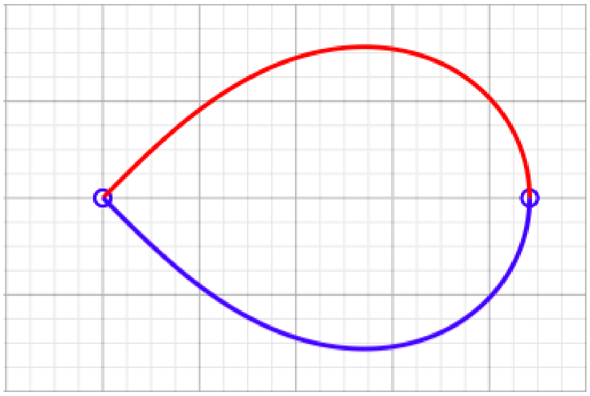
$$y\{8(x^2 + y^2) + 50\} = 0.$$

The quantity within braces can never be zero, since it is always ≥ 50. Hence the only possibility is $y = 0$. Substituting into $2(x^2 + y^2)^2 - 25(x^2 - y^2) = 0$ yields

$$2x^4 = 25x^2$$

i.e. $x = 0$ or $x = \pm\frac{5}{\sqrt{2}}$. These are the points where the graph crosses the x-axis. Look closely at the graph at $(-\frac{5}{\sqrt{2}}, 0)$, $(\frac{5}{\sqrt{2}}, 0)$, and $(0, 0)$. This is where either the tangent is vertical or where there is no tangent (there is no tangent at $(0, 0)$).

It is easy to see (check this!) that if $|x| > \frac{5}{\sqrt{2}}$ then the equation $2(x^2 + y^2)^2 - 25(x^2 - y^2) = 0$ cannot be solved for y over $\mathbf{R}$. So we concentrate on the interval $J = [-\frac{5}{\sqrt{2}}, \frac{5}{\sqrt{2}}]$. Remove the points $x = 0$, $x = \pm\frac{5}{\sqrt{2}}$ from J, since $\frac{\partial\varphi}{\partial y} = 0$ is zero at these places. We are left with the union of the disjoint intervals $I_- = (-\frac{5}{\sqrt{2}}, 0)$ and $I = (\frac{5}{\sqrt{2}}, 0)$. From the graph it is clear that over each of the these intervals, the graph breaks up into two disjoint pieces, each piece looking like the graph of a function. Here are the two disjoint graphs on I.[1] The red part gives one implicit solution, and the blue part another.



To go further, let us consider $a = 0.5$. Solving for y one gets $y = \pm\theta$ where θ is approximately 0.48011, accurate up to four decimal places and rounded off at the fifth. If in the implicit function theorem, we take $b = \theta$, then the graph of the implicit solution $f \colon I \to \mathbf{R}$ with $f(a) = b$ is the red locus. (If we take $b = -\theta$, then the implicit solution satisfying $f(a) = b$ had the blue locus as its graph.) Implicit differentiation is the same as using the chain rule on $\varphi(x, f(x))$ and noting that $\varphi(x, f(x)) \equiv 0$ on I. In greater detail, the chain rule gives

$$\begin{bmatrix} \frac{\partial\varphi}{\partial x}\big|_{(x,f(x))} & \frac{\partial\varphi}{\partial y}\big|_{(x,f(x))} \end{bmatrix} \begin{bmatrix} 1 \\ f'(x) \end{bmatrix} = 0.$$

This amounts to saying that

$$f'(x) = -\frac{\frac{\partial\varphi}{\partial x}\big|_{(x,f(x))}}{\frac{\partial\varphi}{\partial y}\big|_{(x,f(x))}}.$$

---

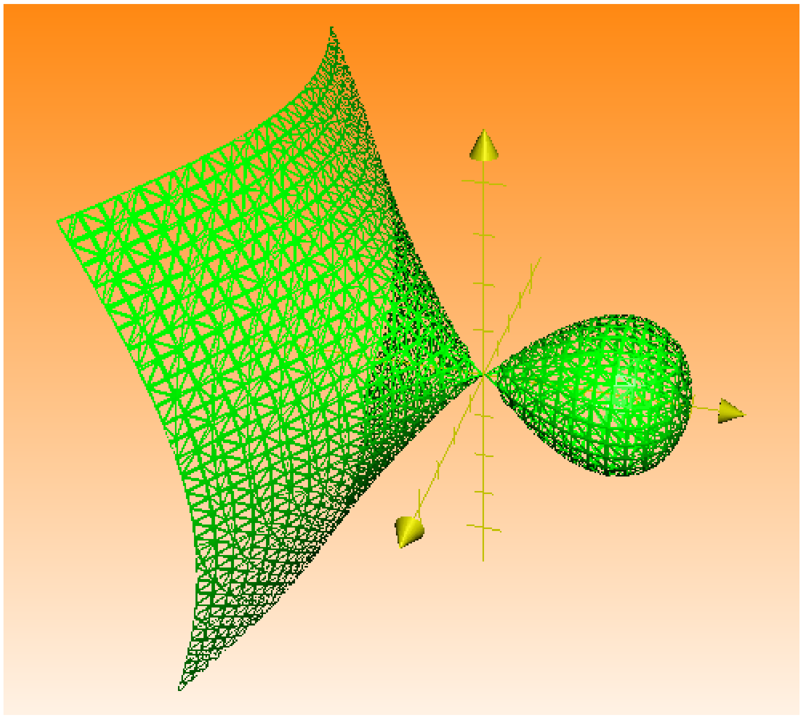
[1] The mirror image about the y-axis will give two disjoint graphs over I_-.

You will get exactly the same answer if you use implicit differentiation from high school. The difference is that now you know a proof that the implicit solution f is differentiable on I, something most of you assumed without proof in high school. In our special case, we get

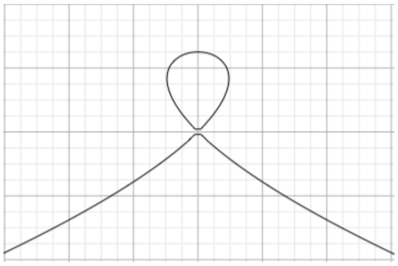
$$y' = \frac{50x - 8x(x^2+y^2)}{8y(x^2+y^2)+50y}.$$

Using this one can work out the slope of the tangent to (a, θ) and it is approximately 0.89273. Here is the picture illustrating this. Our interest is only in the red half, for that is the implicit function such that $f(a) = \theta$.

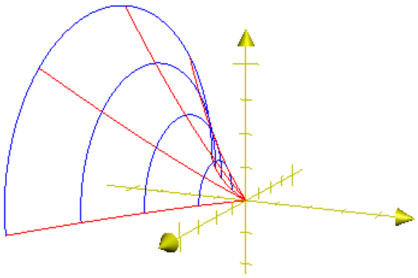
2. Consider the equation $x^2 + y^3 - y^2 + z^2 = 0$. This is the equation of a surface S whose graph is below.



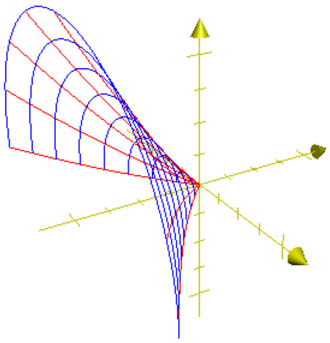
The partial derivative of $\varphi(x, y, z) = x^2 + y^3 - y^2 + z^2$ with respect to z vanishes when $z = 0$. Let C be the intersection of the surface with the xy-plane. In other words, let C be the curve on the xy-plane obtained by setting $z = 0$. Then the equation of C is $x^2 + y^3 - y^2 = 0$. Outside C it is possible to solve z in terms of x and y, according to the implicit function theorem. In fact the solution is a $\mathscr{C}^1$ function of x and y. In each connected region of $\mathbf{R}^2 \setminus C$, there are two implicit solutions, at least according to the picture. Namely the part above the xy-plane, and the part below the plane. Here is the graph of C:



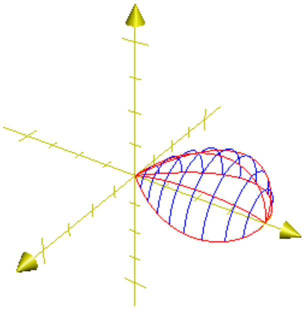
Removing C from the xy-plane leaves us with three connected regions, two are unbounded regions, and one (the one inside the loop) is bounded. If (x, y) lies in the unbounded region above the curve above, then there is no solution for z in terms of (x, y). Over the region lying below the "inverted V-shaped" sub-curve, the surface breaks up into two parts one above the xy-plane and one below the xy-plane, and each piece is a graph of a $\mathscr{C}^1$ implicit solution of z as a function of (x, y). Here is the graph of the implicit solution $z(x, y)$ of the given equation over the region bounded by the V-shaped region lying above the xy-plane. Its reflection about the xy-plane gives another implicit solution.



Here is another view, making the V-shaped region clearer.

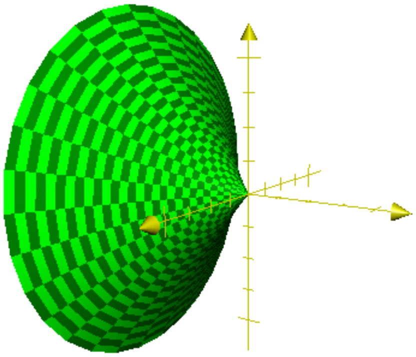


Simlarly, over the loop, the xy-plane bisects S and it falls into two parts, each a graph of an implicit solution. Here is the top part:



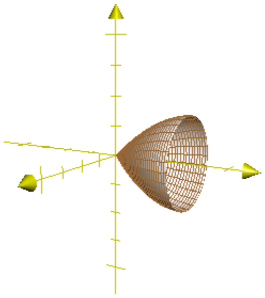
3. Continuing with the above example, one can also try to solve x as a $\mathscr{C}^1$ function of y and z. Since the equation is symmetric with respect to x and z, our analysis for z as a function of (x, y) can be redone by simply exchanging z and x and using the yz-plane instead of the xy-plane as the domain of definition for our functions.

4. Continuing with S and φ as in **2**., what about y as an implicit $\mathscr{C}^1$ function of x and z? This is more interesting. Set $\frac{\partial\varphi}{\partial y}$ equal to 0. We get $3y^2 - 2y = 0$, which means $y = 0$ or $y = \frac{2}{3}$. If one sets $y = 0$ in the given equation of the surface, one gets $x = y = 0$. If one sets $y = \frac{2}{3}$ in the equation of the surface one gets $x^2 + z^2 = \frac{4}{27}$ which is a circle in the xz-plane. Consider the point $(2\sqrt{3}, -2, 0)$ in S. Since $\frac{\partial\varphi}{\partial y}(2\sqrt{3}, -2, 0) \neq 0$, there is an implicit solution of x in terms of y and z in a neighbourhood of $(2\sqrt{3}, 0)$ in the xz-plane. In fact the neighbourhood is the xz-plane punctured at the origin. The resulting graph is



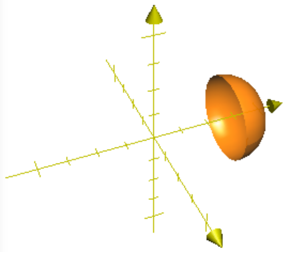
Note that for each point in the xz-plane we have a unique value of y from the above picture, making y a function of x and z. So what was the need to puncture the xz-plane at the origin? Notice the sharp point on the surface when $(x, z) = (0, 0)$. The function is not differentiable there, i.e., $y(x, z)$ is not $\mathscr{C}^1$ if we include $(0, 0)$ in the domain. Upon puncturing, we do get a smooth surface.

The point $(\sqrt{\frac{2}{27}}, \frac{1}{3}, 0)$ on S gives us a diferent implicit $\mathscr{C}^1$ solution of y in terms of x and z and the corresponding graph is:



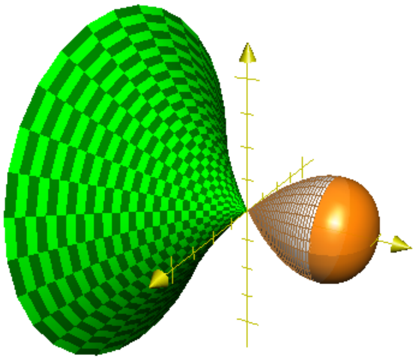
Once again, one eliminates $(0,0)$ because of the sharp edge. One also cannot have (x,z) on the circle $x^2+z^2=\frac{4}{27}$ on the xz-plane plane, since if (x,z) is on the circle, the only points on the connected part of $S\smallsetminus\{\mathbf{0}\}$ containing $(\sqrt{\frac{2}{27}},\frac{1}{3},0)$ and whose projection to the xz-plane lies on the circle, are points with y-coordinate $\frac{2}{3}$ and at these points $\frac{\partial\varphi}{\partial y}$ vanishes. The circle you see as the right most edge of the above surface at the intersection of the cylinder $x^2+z^2=\frac{4}{27}$ with the plane $y=\frac{2}{3}$.

The point $(0,1,0)$ is also a point on S and $\frac{\partial\varphi}{\partial y}(0,1,0)\neq 0$. And the corresponding graph of the implicit solution of y in terms of x and z is



The circle on the left most boundary of the surface is again the intersection of the circular cylinder $x^2+z^2=\frac{4}{27}$ with the plane $y=\frac{2}{3}$.

Here is how you can reconstitute S and see how the three implicit solutions of y in terms of x and z come together:



Compare this with the original picture for S in 2. above.

About these notes. These course notes are a reasonably faithful record of the lectures given at the Chennai Mathematical Institute (CMI) in the January-April semester of 2019-20. The course is Analysis II, a core course for first year undergraduates at CMI. For more material related to this course, visit `https://www.cmi.ac.in/~pramath/teaching.html#ANA2`. If you have comments on these notes, or on related course material, please send an email to pramath@cmi.ac.in.

References

[R] W. Rudin, *Principles of Mathematical Analysis*, (Third Edition), McGraw-Hill, New Delhi, 1976.

[S] M. Spivak, *Calculus on Manifolds (A modern approach to classical theorems of Advanced Calculus)*, Addison-Wesley, Reading, Massachusetts, 1965.