13. Implicit functions

Consider the curve $y^2 = x$ in the plane $\mathbb{R}^2$,

$$C = \{\, (x, y) \in \mathbb{R}^2 \,|\, y^2 = x \,\}.$$

This is not the graph of a function, and yet it is quite close to the graph of a function.

Given any point on the graph, let's say the point $(2, 4)$, we can always find open intervals U containing 2 and V containing 4 and a smooth function $f\colon U \longrightarrow V$ such that $C \cap (U \times V)$ is the graph of f.

Indeed, take $U = (0, \infty)$, $V = (0, \infty)$ and $f(x) = \sqrt{x}$. In fact, we can do this for any point on the graph, apart from the origin. If it is above the x-axis, the function above works. If the point we are interested in is below the x-axis, replace V by $(0, -\infty)$ and $f(x) = \sqrt{x}$, by $g(x) = -\sqrt{x}$.

How can we tell that the origin is a point where we cannot define an implicit function? Well away from the origin, the tangent line is not vertical but at the origin the tangent line is vertical. In other words, if we consider

$$F\colon \mathbb{R}^2 \longrightarrow \mathbb{R},$$

given by $F(x, y) = y^2 - x$, so that C is the set of points where F is zero, then

$$DF(x, y) = (-1, 2y).$$

The locus where we run into trouble, is where $2y = 0$. Somewhat amazingly this works in general:

Theorem 13.1 (Implicit Function Theorem)**.** *Let $A \subset \mathbb{R}^{n+m}$ be an open subset and let $F\colon A \longrightarrow \mathbb{R}^m$ be a $\mathcal{C}^1$-function. Suppose that*

$$(\vec{a}, \vec{b}) \in S = \{\, (\vec{x}, \vec{y}) \in A \,|\, F(\vec{x}, \vec{y}) = \vec{0} \,\}.$$

Assume that

$$\det\left(\frac{\partial F_i}{\partial y_j}\right) \neq 0.$$

Then we may find open subsets $\vec{a} \in U \subset \mathbb{R}^n$ and $\vec{b} \in V \subset \mathbb{R}^m$, where $U \times V \subset A$ and a function $f\colon U \longrightarrow V$ such that $S \cap (U \times V)$ is the graph of f, that is,

$$F(\vec{x}, \vec{y}) = \vec{0} \qquad \text{if and only if} \qquad \vec{y} = f(\vec{x}).$$

where $\vec{x} \in U$ and $\vec{y} \in V$.

Let's look at an example. Let

$$F\colon \mathbb{R}^3 \longrightarrow \mathbb{R},$$

be the function

$$F(x_1, x_2, y) = x_1^3 x_2 - x_2 y^2 + y^5 + 1.$$

Let

$$S = \{ (x_1, x_2, y) \in \mathbb{R}^3 \mid F(x_1, x_2, y) = 0 \}.$$

Then $(1, 3, -1) \in S$. Let's compute the partial derivatives of F,

$$\begin{aligned}
\frac{\partial F}{\partial x_1}(1, 3, -1) &= 3x_1^2 x_2\Big|_{(1,3,-1)} = 9 \\
\frac{\partial F}{\partial x_2}(1, 3, -1) &= (x_1^3 - y^2)\Big|_{(1,3,-1)} = 0 \\
\frac{\partial F}{\partial y}(1, 3, -1) &= (-2x_2 y + 5y^4)\Big|_{(1,3,-1)} = 11.
\end{aligned}$$

So

$$DF(1, 3, -1) = (9, 0, 11).$$

Now what is important is that the last entry is non-zero (so that the 1×1 matrix (1) is invertible). It follows that we may find open subsets $(1, 3) \in U \subset \mathbb{R}^2$ and $-1 \in V \subset \mathbb{R}$ and a $\mathcal{C}^1$ function $f \colon U \longrightarrow V$ such that

$$F(x_1, x_2, f(x_1, x_2)) = 0.$$

It is not possible to write down an explicit formula for f, but we can calculate the partial derivatives of f.

Define a function

$$G \colon U \longrightarrow \mathbb{R},$$

by the rule

$$G(x_1, x_2) = F(x_1, x_2, f(x_1, x_2)) = 0.$$

On the one hand,

$$\frac{\partial G}{\partial x_1} = 0 \qquad \text{and} \qquad \frac{\partial G}{\partial x_2} = 0.$$

On the other hand, by the chain rule,

$$\frac{\partial G}{\partial x_1} = \frac{\partial F}{\partial x_1}\frac{\partial x_1}{\partial x_1} + \frac{\partial F}{\partial x_2}\frac{\partial x_2}{\partial x_1} + \frac{\partial F}{\partial x_3}\frac{\partial f}{\partial x_1}$$

Now

$$\frac{\partial x_1}{\partial x_1} = 1 \qquad \text{and} \qquad \frac{\partial x_2}{\partial x_1} = 0.$$

So

$$\frac{\partial f}{\partial x_1} = -\frac{\frac{\partial F}{\partial x_1}}{\frac{\partial F}{\partial x_3}}.$$

Similarly

$$\frac{\partial f}{\partial x_2} = -\frac{\frac{\partial F}{\partial x_2}}{\frac{\partial F}{\partial x_3}}.$$

So

$$\frac{\partial f}{\partial x_1}(1,3) = -\frac{\frac{\partial F}{\partial x_1}(1,3,-1)}{\frac{\partial F}{\partial x_3}(1,3,-1)} = -\frac{9}{11},$$

and

$$\frac{\partial f}{\partial x_2}(1,3) = -\frac{\frac{\partial F}{\partial x_2}(1,3,-1)}{\frac{\partial F}{\partial x_3}(1,3,-1)} = -\frac{0}{11} = 0.$$

Definition 13.2. *Let $A \subset \mathbb{R}^n$ be an open subset and let $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function.*

*The **directional derivative** of f in the direction of the unit vector $\hat{u}$ is*

$$D_{\hat{u}}f(P) = \lim_{h\to 0} \frac{f(P+h\hat{u}) - f(P)}{h}.$$

If $\hat{u} = \hat{e}_i$ then,

$$D_{\hat{e}_i}f(P) = \frac{\partial f}{\partial x_i}(P),$$

the usual partial derivative.

Proposition 13.3. *If f is differentiable at P then*

$$D_{\hat{u}}f(P) = Df(P)\cdot\hat{u}.$$

Proof. Since A is open, we may find $\delta > 0$ such that the parametrised line

$$r\colon (-\delta,\delta) \longrightarrow A,$$

given by $r(h) = f(P) + h\hat{u}$ is entirely contained in A. Consider the composition of r and f,

$$f\circ r\colon \mathbb{R} \longrightarrow \mathbb{R}.$$

Then

$$\begin{aligned} D_{\hat{u}}f(P) &= \left(\frac{d(f\circ r)}{dh}\right)(0) \\ &= D(r(0))\cdot Dr(0) \\ &= Df(P)\cdot\hat{u}. \end{aligned}$$ □

Note that we can also write

$$D_{\hat{u}}f(P) = \nabla f(P)\cdot\hat{u}.$$

Note that the directional derivative is largest when

$$\hat{u} = \frac{\nabla f(P)}{\|\nabla f(P)\|},$$

so that the gradient always points in the direction of maximal change (and in fact the magnitude of the gradient, gives the maximum change). Note also that the directional derivative is zero if $\hat{u}$ is orthogonal to the gradient and that the directional derivative is smallest when

$$\hat{u} = -\frac{\nabla f(P)}{\|\nabla f(P)\|}.$$

Proposition 13.4. *If $\nabla f(P) \neq 0$ then the tangent hyperplane Π to the hypersurface*

$$S = \{ Q \in \mathbb{R}^n \mid f(Q) - f(P) = 0 \},$$

is the set of all points Q which satisfy the equation

$$\nabla f(P) \cdot \overrightarrow{PQ} = 0.$$

Remark 13.5. *If f is $\mathcal{C}^1$, then f is the graph of some function, locally about P.*

Proof. By definition, the point Q belongs to the tangent hyperplane if and only if there is a curve

$$r\colon (-\delta, \delta) \longrightarrow S,$$

such that

$$r(0) = P \qquad \text{and} \qquad r'(0) = \overrightarrow{PQ}.$$

Now, since $r(h) \in S$ for all $h \in (-\delta, \delta)$, we have $F(r(h)) = 0$. So

$$\begin{aligned} 0 &= \frac{dF(r(h))}{dh}(0) \\ &= \nabla F(r(0)) \cdot r'(0) \\ &= \nabla F(P) \cdot \overrightarrow{PQ}. \end{aligned}$$

□