Definition

Consider a set X, and a distance function $d : X \times X \to \mathbb{R}$ satisfying

1. $d(x, y) = d(y, x)$ (symmetric),
2. $d(x, y) = 0 \;\Leftrightarrow\; x = y$
3. $d(x, y) + d(y, z) \geq d(x, z)$ (triangle inequality)

for all $x, y, z \in X$. Then (X, d) is called a *metric space*.

Note that it follows from the above prescribed conditions, that the distance function is *positive definite*: $d(x, y) \geq 0$.

We define the open r-ball at $x \in X$ as

$$B(x, r) := \{y \in X \mid d(x, y) < r\}.$$

A set $A \subset X$ is called bounded if it is contained in an r-ball for some $r < \infty$, and open if for all $x \in A$ there exists an r such that $B(x, r) \subset A$. The interior of a set is the union of all its open subsets. Any open subset of X containing x is called a neighbourhood of $x \in X$. A point $x \in X$ is a boundary point of a subset $A \subset X$ if for all neighbourhoods U of x, we have $U \cap A \neq \emptyset$ and $U \setminus A \neq \emptyset$. The boundary ∂A of $A \subset X$ is the set of all boundary points of A. The closure of A, is defined as the set

$$\bar{A} := \{x \in X \mid B(x, r) \cap A \neq \emptyset, \ \forall r > 0\}.$$

A set A is closed is $A = \bar{A}$, and $A \subset X$ is dense in X if $\bar{A} = X$. A set A is nowhere dense if its closure has empty interior.

A point $x \in X$ is called an accumulation point of a set $A \subset X$ if all balls $B(x, \varepsilon)$ intersect A. The set of accumulation points of A is called the derived set A'. A set A is closed if $A' \subset A$ and $\bar{A} = A \cup A'$. A is called perfect if $A = A'$.

We say that a sequence $\{x_n\}_{n\in\mathbb{N}}$ converges (as $n \to \infty$) if $\forall \varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\forall n \geq N$ we have $d(x_n, x) < \varepsilon$. We say that two sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ converge exponentially (or with exponential speed) to each other if $d(x_n, y_n) < cd^n$ for some $c > 0$ and $0 \leq d < 1$. The sequence is a Cauchy sequence if $\forall\ \varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_i, x_j) < \varepsilon$ whenever $i, j \geq N$.

A metric space is called complete if every Cauchy sequence converges in it.

Examples of complete metric spaces are $\mathbb{R}^n$ with the (usual) Euclidean metric, and all closed subsets of $\mathbb{R}^n$ with this metric.

Definition (Contraction)

A map $F : X \to X$, where (X, d) is a metric space, is a contraction if there exists $K < 1$ such that

$$d(F(x), F(y)) \leq Kd(x, y), \quad \forall x, y \in X. \tag{10}$$

A condition of the type (10) is called a Lipschitz condition, where $K \geq 0$ is called the Lipschitz constant. Contractions are thus Lipschitz maps with a Lipschitz constant that is smaller than 1.

We now formulate the central result about contractions.

Theorem (Contraction mapping theorem)

Let X be a complete metric space, and $F : X \to X$ be a contraction. Then F has a unique fixed point, and under the action of iterates of $F : X \to X$, all points converge with exponential speed to it.

Iterating $d(F(x), F(y)) \leq Kd(x, y)$ gives

$$d(F^n(x), F^n(y)) \leq K^n d(x, y), \tag{11}$$

with $x, y \in X$ and $n \in \mathbb{N}$. Thus $(F^n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence, because with $m > n$ we have

$$\begin{aligned} d(F^m(x), F^n(x)) &\leq \sum_{k=0}^{m-n-1} d(F^{n+k+1}(x), F^{n+k}(x)) \\ &\leq \sum_{k=0}^{m-n-1} K^{n+k} d(F(x), (x)) \leq \frac{K^n}{1-K} d(F(x), x) \end{aligned}$$

and $K^n \to 0$ as $n \to \infty$. In the last step we used the fact that with $0 \leq K < 1$ it follows that

$$\sum_{k=0}^{m-n-1} K^k \leq \sum_{k=0}^{\infty} K^k = \frac{1}{1-K}.$$

Thus the limit $\lim_{n\to\infty} F^n(x)$ exists because Cauchy sequences converge in X. We denote the limit x_0.

By (11) under iteration by F all points in X converge to the same point as $\lim_{n\to\infty} d(F^n(x), F^n(y)) = 0$ for all $x, y \in X$ so that if x converges to x_0 then so does any $y \in X$.
It remains to be shown that x_0 is a fixed point of F: $F(x_0) = x_0$.
Using the triangle inequality we have

$$\begin{aligned} d(x_0, F(x_0)) &\leq d(x_0, F^n(x)) + d(F^n(x), F^{n+1}(x)) \\ &\quad + d(F^{n+1}(x), F(x_0)) \\ &\leq (1+K) d(x_0, F^n(x)) + K^n d(x, F(x)), \end{aligned}$$

for all $x \in X$ and $n \in \mathbb{N}$. The right-hand-side of this inequality tends to zero as $n \to \infty$, and hence $F(x_0) = x_0$.

Proposition

Let $I \subset \mathbb{R}$ is a closed bounded interval, and $F : I \to I$ a continuously differentiable (C^1) function with $|F'(x)| < 1$ for all $x \in I$. Then F is a contraction.

Proof.

First we show that if $|F'(x)| \leq K$ then F is Lipschitz with Lipschitz constant K. By the Mean Value Theorem, for any two points $x, y \in I$ there exists a c between x and y such that

$$\begin{aligned} d(F(x), F(y)) &= |F(x) - F(y)| = |F'(c)(x - y)| \\ &= |F'(c)|d(x, y) \leq Kd(x, y). \end{aligned}$$

At some point $x_0 \in I$ the maximum of $|F'(x)|$ will be attained since F is continuous, and $|F'(x_0)| < 1$. □

Example (Fibonacci's rabbits)

Leonardo Pisano, better known as *Fibonacci*, tried to understand how many pairs of rabbits can be grown from one pair in one year. He figured out that each pair breads a pair every month, but a newborn pair only breads in the second month after birth. Let b_n denote the number of rabbit pairs at time n. Let $b_0 = 1$ and in the firts month they bread one pair so $b_1 = 2$. At time $n = 2$, again one pair is bread (from the one that were around at time $n = 1$, the other one does not yet have the required age to bread). Hence, $b_2 = b_1 + b_0$. Subsequently, $b_{n+1} = b_n + b_{n-1}$. Expecting the growth to be exponential we would like to see how fast these number grow, by calculating $a_n = b_{n+1}/b_n$. Namely, if $b_n \to cd^n$ as $n \to \infty$ for some c, d then $b_{n+1}/b_n \to d$.

Example (cont)

We have

$$a_{n+1} = b_{n+2}/b_{n+1} = \frac{1}{a_n} + 1.$$

Thus $\{a_n\}_{n\in\mathbb{N}}$ is the orbit of $a_0 = 1$ of the map $g(x) = 1/x + 1$. We have $g'(x) = -x^{-2}$. Thus g is not a contraction on $(0, \infty)$. But we note that $a_1 = 2$ and consider the map g on the closed interval $[3/2, 2]$. We have $g(3/2) = 5/3 > 3/2$ and $g(2) = 3/2$. Hence $g([3/2, 2]) \subset [3/2, 2]$. Furthermore, fro $x \in [3/2, 2]$ we have $|g'(x)| = 1/x^2 \leq 4/9 < 1$ so that g is a contraction on $[3/2, 2]$. Hence, by the contraction mapping theorem, there exists a unique fixed point, so $\lim_{n\to\infty} a_n$ exists. The solution is a fixed point of $g(x)$, yielding $x^2 - x - 1 = 0$. The only positive root of this equation is $x = (1 + \sqrt{5})/2$.

Finding the roots (preimages of zero) of a function $F : \mathbb{R} \to \mathbb{R}$ is difficult in general. Newton's method is an approach to find such roots through iteration. The idea is rather straight forward. Suppose x_0 is a guess for a root. We would like to improve our guess by chosing an improved approximation x_1. We write the first order Taylor expansion of F at x_1 in terms of our knowledge about F at x_0: $F(x_1) = F(x_0) + F'(x_0)(x_1 - x_0)$. By setting $F(x_1) = 0$ (our aim), we obtain from the Taylor expansion that

$$x_1 = x_0 - F(x_0)/F'(x_0) =: G(x_0). \tag{12}$$

We note that a fixed point y of G corresponds to a root of F if $F'(y) \neq 0$. We call a fixed point y of a differentiable map G *superattracting* if $G'(y) = 0$. We have

Proposition

If $|F'(x)| > \delta$ for some $\delta > 0$ and $|F''(x)| < M$ for some $M < \infty$ on a neighbourhood of a root r (satisfying $F(r) = 0$), then r is a superattracting fixed point of G (cf (12)).

Proof.

We observe that $G'(x) = F(x)F''(x)/(F'(x))^2$. Note that G is a contraction on a neighbourhood of r. □

There is a higher dimensional version of this result, which requires us to introduce the notion of the derivative DF of a map $F : \mathbb{R}^m \to \mathbb{R}^m$:

$$DF(x)y = \lim_{\varepsilon \to 0} \frac{F(x + \varepsilon y) - F(x)}{\varepsilon}.$$

Making a Taylor expansion of F in ε, and denoting $F = (F_1, \ldots, F_m)$ where F_i denotes the ith component of the map we obtain

$$F_i(x + \varepsilon y) = F_i(x) + \varepsilon \nabla F_i(x) \cdot y + o(\varepsilon),$$

yielding that $(DF(x)y)_i = \nabla F_i(x) \cdot y$. In other words, DF is a linear map from $\mathbb{R}^m$ to $\mathbb{R}^m$ which we may represent by the so-called *Jacobian matrix*

$$DF(x) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x) & \cdots & \frac{\partial F_1}{\partial x_m}(x) \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1}(x) & \cdots & \frac{\partial F_m}{\partial x_m}(x) \end{pmatrix}.$$

where x_i denotes the ith component of the vector $x = (x_1, \ldots, x_m)$.

For this derivation to be meaningful, we need the first derivative of F_i with respect to x_j for all $i, j = 1, \dots m$ to exist. If one of these does not exist then the map F is not differentiable.
For completeness, we now state the derivative test in $\mathbb{R}^m$ without proof. Recall that a *strictly convex set* $C \subset \mathbb{R}^n$ is a set C such that for all $a, b \in \bar{C}$, the line segment with endpoints a, b is entirely contained in C, except possibly for one or both endpoints. Also, let the norm $||A||$ of a linear map A is defined by $||A|| := \max_{|v|=1} |A(v)|$.

Theorem

If $C \subset \mathbb{R}^n$ is an open strictly convex set, $\bar{C}$ its closure, $F : \bar{C} \to \mathbb{R}^n$ differentiable on C and continuous on $\bar{C}$ with $||DF|| \leq K < 1$ on C, then F has a unique fixed point $x_0 \in \bar{C}$ and $d(F^n(x), x_0) \leq K^n d(x, x_0)$ for every $x \in \bar{C}$.

We note that this result is in agreement with the fact that equilibria of linear autonomous ODEs with all eigenvalues having negative real part are asymptotically stable (with exponential convergence).

The inverse function theorem says that if a differentiable map has invertible derivative at some point, then the map is invertible near that point. It is thus related to "linearizability": if the linearization of a map in a point is invertible, then so is the nonlinear map in a neighbourhood of this point.
We first consider the simplest version of the inverse function theorem, in $\mathbb{R}$.

Theorem (Inverse function theorem in $\mathbb{R}$)

Suppose $I \subset \mathbb{R}$ is an open interval and $F : I \to \mathbb{R}$ is a differentiable function. If a is such that $F'(a) \neq 0$ and F' is continuous at a, then F is invertible on a neighbourhood U of a and for all $x \in U$ we have $(F^{-1})'(y) = 1/F'(x)$, where $y = F(x)$.

NB: if F is C^r then it can be shown that F^{-1} is C^r as well.

Proof: The proof is by application of the contraction mapping theorem. We consider the map

$$\phi_y(x) = x + \frac{y - F(x)}{F'(a)}$$

on I. Fixed points of ϕ_y are solutions of our problem since $\phi_y(x) = x$ if and only if $F(x) = y$.

We now show that ϕ_y is a contraction in some closed neighbourhood of $a \in I$. Then by the contraction mapping theorem, ϕ_y has a unique fixed point, and hence there exists a unique x such that $F(x) = y$ for y close enough to $F(a)$.

Let $A = F'(a)$ and $\alpha := |A|/2$. By continuity of F' at a there is an $\varepsilon > 0$ such that with $W := (a - \varepsilon, a + \varepsilon) \subset I$ we have $|F'(x) - A| < \alpha$ for x in the closure $\bar{W}$ of W.

To see that ϕ_y is a contraction on $\bar{W}$ we observe that if $x \in \bar{W}$ we have

$$|\phi_y'(x)| = \left|1 - \frac{F'(x)}{A}\right| = \left|\frac{A - F'(x)}{A}\right| < \frac{\alpha}{|A|} = 1/2.$$

Now, using Proposition 7 we obtain $|\phi_y(x) - \phi_y(x')| \le |x - x'|/2$ for all $x, x' \in \bar{W}$.

We also need to show that $\phi_y(\bar{W}) \subset \bar{W}$ for y suffiently close to $b := F(a)$. Let $\delta = |A|\varepsilon/2$ and $V = (b - \delta, b + \delta)$. Then for $y \in V$ we have

$$|\phi_y(a) - a| = \left|a - \frac{y - F(a)}{A} - a\right| = \left|\frac{y - b}{A}\right| < \left|\frac{\delta}{A}\right| = \frac{\varepsilon}{2}.$$

So if $x \in \bar{W}$ then

$$|\phi_y(x) - a| \le |\phi_y(x) - \phi_y(a)| + |\phi_y(a) - a| \le \frac{|x - a|}{2} + \frac{\varepsilon}{2} \le \varepsilon,$$

and hence $\phi_y(x) \in \bar{W}$.

Hence, if $y \in V$ then $\phi_y : \bar{W} \to \bar{W}$ has a unique fixed point $G(y) \in W$ which depends continuously on y.

Next we prove that the inverse is differentiable: for $y = F(x) \in V$ we will show that $G'(y) = 1/B$ where $B := F'(G(y))$.
Let $U := G(V) = W \cap F^{-1}(V)$, which is open. Take $y + k = F(x + h) \in V$. Then

$$\frac{|h|}{2} \geq |\phi_y(x+h) - \phi_y(x)| = \left| h + \frac{F(x) - F(x+h)}{A} \right| = \left| h - \frac{k}{A} \right| \geq |h| - |K/A|.$$

Hence, we have

$$\frac{|h|}{2} \leq \left| \frac{k}{A} \right| < \frac{|k|}{\alpha} \quad \text{and} \quad \frac{1}{|k|} < \frac{2}{\alpha |h|}.$$

Since
$G(y + k) - G(y) - k/B = h - k/B = -(F(x + h) - F(x) - Bh)/B$ we obtain

$$\frac{|G(y+k) - G(y) - k/B|}{|k|} < \frac{2}{|B|\alpha} \frac{|F(y+h) - F(y) - Bh|}{|h|} \to 0 \text{ as } |h| \leq |k|/\alpha$$

This proves that $G'(y) = 1/B$.

Example

Let $F(x) = \sin(x)$. W have $F'(0) = 1$. Hence, F is invertible near 0. Being assured of the fact that the inverse locally exists, it makes sense to derive a Taylor expansion of it. Let $G = F^{-1}$ be define in a small neighbourhood of $F(0) = 0$, where it satisfies $G(\sin(x)) = x$. We obtain a Taylor expansion of G by substituting the Taylor expansion of $\sin(x)$ and that of G in this equation and resolve the equaliy at each order in x. We write $G(y) = ay + by^2 + cy^3 + dy^4 + O(y^5)$ and $\sin(x) = x - \frac{1}{6}x^3 + O(x^5)$. Matching Taylor coefficients we obtain $a = b = 1$, $c = \frac{1}{6}$ and $d = 0$ so that

$$F^{-1}(x) = x + \frac{1}{6}x^3 + O(x^5).$$

Without too much difficulty (replacing some numbers by linear maps and some absolute values by matrix norms) a similar result can be proven for maps of $\mathbb{R}^m$. (We leave this as an exercise.)

Theorem (Inverse function theorem in $\mathbb{R}^m$)

Suppose $O \subset \mathbb{R}^m$ is open, $F : O \to \mathbb{R}^m$ differentiable, and DF is invertible at a point $\mathbf{a} \in O$ and continuous at $\mathbf{a}$. Then there exist neighbourhoods $U \subset O$ of $\mathbf{a}$ and V of $\mathbf{b} := F(\mathbf{a}) \in \mathbb{R}^m$ such that F is a bijection from U to V [i.e. F is one-to-one on U and $F(U) = V$]. The inverse $G : V \to U$ of F is differentiable with $DG(y) = (DF(G(y)))^{-1}$. Furthermore, if F is C^r on U, then so is its inverse (on V).

Example

Consider the map $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 - y \\ -x \end{pmatrix}.$$

Then

$$DF(x) = \begin{pmatrix} 2x & -1 \\ -1 & 0 \end{pmatrix},$$

from which it follows hat $DF(x)$ is invertible for all x since $\det(DF(x)) = -1$, and noninvertibility would require that his determinant is equal to zero. The fact that the derivative is invertible for all $x \in \mathbb{R}^2$ appears to imply that F is invertible on all of $\mathbb{R}^2$. And indeed, the inverse of F can be computed to be

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ y^2 - x \end{pmatrix}.$$

The Implicit Function Theorem (IFT) establishes, under the assumption of some conditions on derivatives, that if we can solve an equation for a particular parameter value, then there is a solution for nearby parameters as well. We illustrate the principle with a linear map $A : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^m$. We write $A := (A_1, A_2)$, where $A_1 : \mathbb{R}^m \to \mathbb{R}^m$ and $A_2 : \mathbb{R}^p \to \mathbb{R}^m$ are linear. Suppose we pick $y \in \mathbb{R}^p$ and want to find $x \in \mathbb{R}^m$ so that $A(x, y) = 0$. To see when this can be done, write $A_1 x + A_2 y = 0$ as

$$A(x, y) = 0 \quad \Leftrightarrow \quad x = -(A_1)^{-1} A_2 y := Ly. \tag{13}$$

We can interpret this as saying that $A(x, y) = 0$ implicitly defines a map $L : \mathbb{R}^p \to \mathbb{R}^m$ such that $A(Ly, y) = 0$. The crucial condition transpiring from this manipulation is that A_1 needs to be invertible.

The IFT asserts that this property naturally extends to nonlinear maps $F : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^m$, in the neighbourhood of a point $(\mathbf{a}, \mathbf{b})$ where $F(\mathbf{a}, \mathbf{b}) = 0$, the corresponding condition being that $D_1 F(\mathbf{a}, \mathbf{b})$ (denoting the derivative with respect to the first variable) is invertible. The IFT is closely related to the Inverse Function Theorem, and can be derived directly from it.

Theorem (Implicit Function Theorem in $\mathbb{R}^m$)

Let $O \subset \mathbb{R}^m \times \mathbb{R}^p$ be open and $F : O \to \mathbb{R}^m$ a C^r map. If there is a point $(\mathbf{a}, \mathbf{b}) \in O$ such that $F(\mathbf{a}, \mathbf{b}) = 0$ and $D_1 F(\mathbf{a}, \mathbf{b})$ is invertible, then there are open neighbourhoods $U \subset O$ of $(\mathbf{a}, \mathbf{b})$ and $V \subset \mathbb{R}^p$ of $\mathbf{b}$ such that for every $y \in V$ there exists a unique $x =: G(y) \in \mathbb{R}^m$ with $(x, y) \in U$ and $F(x, y) = 0$. Furthermore, G is C^r and $DG(y) = -(D_1 F(x, y))^{-1} D_2 F(x, y)$.

Proof.

The map $H(x, y) := (F(x, y), y) : O \to \mathbb{R}^m \times \mathbb{R}^p$ is C^r then $DH(\mathbf{a}, \mathbf{b})(x, y) = (D_1 F(\mathbf{a}, \mathbf{b})x + D_2 F(\mathbf{a}, \mathbf{b})y, y)$. This is equal to $(0, 0)$ only if $y = 0$ and $D_1 F(\mathbf{a}, \mathbf{b})x = 0$, which implies that $x = 0$ if $D_1 F(\mathbf{a}, \mathbf{b})$ is invertible. Hence DH is invertible and by the Inverse Function Theorem there are open neighbourhoods $U \subset O$ of $(\mathbf{a}, \mathbf{b})$ and $W \subset \mathbb{R}^m \times \mathbb{R}^p$ of $(0, \mathbf{b})$ such that $H : U \to W$ is invertible with C^r inverse $H^{-1} : W \to U$. Thus, for any $y \in V := \{y \in \mathbb{R}^p \mid (0, y) \in W\}$ there exists an $x := G(y) \in \mathbb{R}^m$ such that $(x, y) \in U$ and $H(x, y) = (0, y)$, or equivalently $F(x, y) = 0$.
Now $(G(y), y) = (x, y) = H^{-1}(0, y)$ and hence G is C^r. To find $DG(\mathbf{b})$, let $\gamma(y) := (G(y), y)$. Then $F(\gamma(y)) = 0$ and hence $DF(\gamma(y))D\gamma(y) = 0$ by the chain rule. For $y = \mathbf{b}$ this gives $D_1 F(\mathbf{a}, \mathbf{b})DG(\mathbf{b}) + D_2 F(\mathbf{a}, \mathbf{b}) = DF(\mathbf{a}, \mathbf{b})D\gamma(\mathbf{b}) = 0$, completing the proof. □

Example

Let $F : \mathbb{R} \to \mathbb{R}$ where $F(x, \lambda) = \sin(x) + \lambda$ we know that $F(0, 0) = 0$ and would like to know about the existence of roots near $x = 0$ is λ is small. Since $D_1 F(0, 0) = 1 \neq 0$ the IFT asserts that if λ is small, there exists a unique $x(\lambda)$ near 0 such that $F(x(\lambda)) = 0$.

Persistence of transverse intersections

Consider two curves in the plane $\mathbb{R}^2$. Let they have the parametrized form $f, g : \mathbb{R} \to \mathbb{R}^2$. Then the intersection points of these curves are roots of the equation $h : \mathbb{R}^2 \to \mathbb{R}^2$ with $h(s,t) = f(s) - g(t)$. Suppose they have an intersection at $f() = g(t)$ with $(s,t) = (0,0)$. Writing $f(s) = (f_1(s), f_2(s))^T$ and $g(s) = (g_1(s), g_2(s))^T$ we obtain

$$Dh = \begin{pmatrix} \frac{df_1}{ds}(0) & -\frac{dg_1}{dt}(0) \\ \frac{df_2}{ds}(0) & -\frac{dg_2}{dt}(0) \end{pmatrix}.$$

The first column vector is the tangent vector to the curve of f and the second vector is the tangent vector to the curve of g. Namely, thinking of the tangent as the best linear approximation to the curve, we find

$$f(s) = f(0) + s\frac{df}{ds}(0) + O(s^2).$$

so that indeed $\frac{df}{ds} = (\frac{df_2}{ds}(0), \frac{df_2}{ds}(0))$ is the tangent vector at $s = 0$. Suppose now that the curves depend smoothly on some parameter $\lambda \in \mathbb{R}$, yielding parametrizations f_λ and g_λ, then the intersections are given by roots of $h_\lambda = f_\lambda - g_\lambda$.

Assume that at $\lambda = 0$ there is an intersection of the curves at $(s, t) = (0, 0)$. We would like to understand what happens if λ is perturbed away from 0. It follows from the IFT that if $h_0(0, 0) = 0$ and $Dh_0(0, 0)$ is nonsingular, for sufficiently small λ, there exists unique smooth functions $s(\lambda)$ and $t(\lambda)$ so that $h_\lambda(s(\lambda), t(\lambda)) = 0$ near $(0, 0)$. We refer to this locally smooth variation of the intersection point as persistence.

The condition that $Dh_0(0, 0)$ is nonsingular is related to transversality. We call the linear subspace generated by the tangent vector to the curve for f transversal to the linear subspace generated by the tangent vector to the curve for g if these tangent vectors span $\mathbb{R}^2$. The latter depends on the fact whether these vectors are linearly independent, which is identical to the nonsingularity condition that $\det(Dh) \neq 0$. We call the intersection of the two curves transverse if the corresponding tangent vectors span the $\mathbb{R}^2$.

We thus find that transverse intersections of curves in the plane are persistent. This is an illustration of a more general theorem stating that transverse intersections are persistent. It actually turns out that *typically* intersections of curves are transverse.

We note that the Inverse and Implicit Function Theorems can be proven not only in $\mathbb{R}^m$ but also in more general *Banach spaces* (which are complete normed vector spaces). There are any important examples of (infinite dimensional) function spaces that are Banach spaces.