LECTURE: INVERSE AND IMPLICIT FUNCTION THEOREMS

1. Inverse Function Theorem

Goal: If $y = f(x)$, when can we solve for x in terms of y? That is, when can we write $x = g(y)$ where g is a **smooth** function?

Intuitively, we would like to say: As long as $f'(x) \neq 0$, we can solve for x in terms of y, at least locally, and moreover

$$g'(x) = \frac{1}{f'(g(x))}$$

Which follows by differentiating the equation $f(g(x)) = x$

Inverse Function Theorem:

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be C^1 and suppose $DF(x_0)$ is invertible at x_0.

Then F is invertible in a neighborhood of x_0. More precisely:

(1) There exist neighborhoods U of x_0 and V of $y_0 = F(x_0)$ such that the restriction $F|_U : U \to V$ is a bijection.

(2) The inverse function $G : V \to U$ is also continuously differentiable, and for $y \in V$

$$DG(y) = [DF(G(y))]^{-1}$$

Proof-Outline:

WLOG, assume $x_0 = 0$

STEP 1: Since $DF(0)^{-1}DF(x)$ is continuous at $x = 0$ and $DF(0)^{-1}DF(0) = I$ we can find $\delta > 0$ such that for all $\|x\| \leq \delta$, $DF(x)$ is invertible and

$$\|I - DF(0)^{-1}DF(x)\| \leq \frac{1}{2}$$

STEP 2: Let $B = \overline{B(0, \delta)}$

Given a parameter y define the "Newton Map"

$$N(x; y) = x - DF(0)^{-1}(F(x) - y).$$

Notice x is a fixed point of $N(\cdot; y)$ if and only if $y = F(x)$

STEP 3: We just need to verify the hypotheses of the Banach fixed point theorem, that is show (will be skipped)

(1) $N(\cdot; y) : B \to B$ for all y in a neighborhood V of $f(0)$.

(2) $N(\cdot; y)$ is contraction on B.

STEP 4: Then, for all $y \in V$, use the Banach fixed point theorem to find a unique $x \in B$ such that $f(x) = y$. Let $f^{-1}(y)$ be this unique x.

STEP 4: Show $f^{-1}(y)$ is continuous and differentiable on V. This can be done directly but it'll be much easier once we prove the Uniform Contraction Mapping Principle below. $\square$

2. Implicit Function Theorem

On the other side of the coin is the Implicit Function Theorem.

Goal: Suppose you have an equation of the form $F(x,y) = 0$, can you solve for one variable in terms of the other one(s)?

Example 1: Let $F(x,y) = x^2 + y^2 - 1 = 0$, that is $x^2 + y^2 = 1$. Then you can solve for y in terms of x because $y = \pm\sqrt{1-x^2}$. This expression fails precisely when $y = 0$ that is when $F_y = 0$ (this is the derivative of F with respect to the variable you want to solve for)

Moreover, we can calculate $\frac{dy}{dx}$ in terms of partial derivatives:

$$\begin{aligned}\left(x^2 + y^2 - 1\right)' &= (0)' \\ 2x + 2y\frac{dy}{dx} &= 0 \\ \frac{dy}{dx} &= -\frac{2x}{2y} \\ \frac{dy}{dx} &= -\frac{F_x}{F_y}\end{aligned}$$

Notice how the x and y get switched in the right-hand-side. Again, notice how this is defined when $F_y \neq 0$

Example 2: To prep for the notation of the Implicit Function Thm:

Let $n = 3$ (number of x variables) and $m = 2$ (number of y variables), and define

$$F : \mathbb{R}^{3+2} \to \mathbb{R}^2 \text{ by } F = (F_1, F_2) \text{ where}$$

$$\begin{aligned} F_1(x_1, x_2, x_3, y_1, y_2) =& x_1 y_2 - 4x_2 + 3 + 2e^{y_1} \\ F_2(x_1, x_2, x_3, y_1, y_2) =& 2x_1 - x_3 + y_2 \cos(y_1) - 6y_1 \end{aligned}$$

Notice $F(x_0, y_0) = 0$ where $x_0 = (3, 2, 7)$ and $y_0 = (0, 1)$

Question: Can we solve for y in terms of x, for x near $x_0 = (3, 2, 7)$?

The implicit function theorem says yes provided that "$F_y \neq 0$" (the derivative with respect to the variable you want to solve for is nonzero)

$$[F'(x, y)] = \left[\begin{bmatrix} y_2 & -4 & 0 \\ 2 & 0 & -1 \end{bmatrix} \quad | \quad \begin{bmatrix} 2e^{y_1} & x_1 \\ -y_2 \sin(y_1) - 6 & \cos(y_1) \end{bmatrix} \right] = \begin{bmatrix} F_x & | & F_y \end{bmatrix}$$

$$[F'(x_0, y_0)] = \left[\begin{bmatrix} 1 & -4 & 0 \\ 2 & 0 & -1 \end{bmatrix} \quad | \quad \begin{bmatrix} 2 & 3 \\ -6 & 1 \end{bmatrix} \right] = \begin{bmatrix} F_x(x_0, y_0) & | & F_y(x_0, y_0) \end{bmatrix}$$

Here all you need to check here is that $F_y(x_0, y_0)$ is invertible, but

$$|F_y(x_0, y_0)| = \begin{vmatrix} 2 & 3 \\ -6 & 1 \end{vmatrix} = 2 + 18 = 20 \neq 0 \text{ YES}$$

Then the implicit function theorem then that there is a function $y = G(x)$ from a neighborhood W of $x_0 = (3, 2, 7)$ (x variables) to $\mathbb{R}^m$ such that $F(x, G(x)) = 0$ (the equation is satisfied)

Moreover, we can calculate $G'(3, 2, 7)$ via

$$G'(3, 2, 7) = -\left(F_y\right)^{-1} F_x = -\begin{bmatrix} 2 & 3 \\ -6 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & -4 & 0 \\ 2 & 0 & -1 \end{bmatrix} = -\frac{1}{20} \begin{bmatrix} 5 & 4 & -3 \\ -10 & 12 & 2 \end{bmatrix}$$

Looking at the $(1, 2)$ entry for example, this tells us $\frac{\partial y_1}{\partial x_2} = -\frac{4}{20} = -\frac{1}{5}$

Compare this once again with the $\frac{dy}{dx} = -\frac{F_x}{F_y}$ condition from Example 1.

Implicit Function Theorem:

Suppose $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$ is C^1 and $F(x_0, y_0) = 0$ for some (x_0, y_0)

If $\det F_y(x_0, y_0) \neq 0$, then there is an open neighborhood U of (x_0, y_0) and an open neighborhood W of x_0 and a function $G : W \to \mathbb{R}^m$ differentiable at x_0 such that

$$\{(x, y) \in U \mid F(x, y) = 0\} = \{(x, G(x)) \mid x \in W\}$$

$$\text{Moreover } G'(x_0) = -\left(F_y(x_0, y_0)\right)^{-1} F_x(x_0, y_0)$$

In other words, if the derivative with respect to the variable you want to solve for is invertible, then the equation $F(x, y) = 0$ is locally the graph of a function $y = G(x)$.

Application: This theorem is extremely useful in PDEs. Lots of PDEs, especially first-order ones, are usually given by implicit equations of the form $F(x, u, \nabla u) = 0$. The implicit function theorem can then be used to solve for u in terms of x, provided some "nondegeneracy" condition holds, which is usually equivalent to the assumption above.

Proof:[1]

Surprisingly, the Implicit function theorem and Inverse function theorem are equivalent (notice they both solve for one variable in terms of another one), so our goal is to apply the Inverse Function Theorem to a cleverly designed function

[1] The proof is taken from this website

STEP 1: Given our F, define $f : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ by

$$f(x, y) = (x, F(x, y))$$

Goal: Apply the Inverse function theorem to f at $a = (x_0, y_0)$

First show that $f'(a)$ is invertible. However

$$[f'(x, y)] = \begin{bmatrix} I_{n\times n} & 0_{n\times m} \\ F_x & F_y \end{bmatrix}$$

$$\text{Hence } [f'(a)] = \begin{bmatrix} I_{n\times n} & 0_{n\times m} \\ F_x(x_0, y_0) & F_y(x_0, y_0) \end{bmatrix}$$

It then follows from cofactor expansion along the first n rows that

$$\det[f'(a)] = \det[F_y(x_0, y_0)] \neq 0$$

Where the last step follows precisely because $F_y(x_0, y_0)$ is invertible

STEP 2: Hence $f'(a)$ is invertible and therefore by the Inverse Function Theorem there is an open set U containing (x_0, y_0) and an open set V containing $f(x_0, y_0)$ such that $f : U \to V$ is invertible.

Moreover, $f^{-1} : V \to U$ is differentiable at $f(x_0, y_0)$ and

$$\left(f^{-1}\right)' (f(x_0, y_0)) = (f'(x_0, y_0))^{-1}$$

Note: $f(x_0, y_0) = (x_0, F(x_0, y_0)) = (x_0, 0)$

Write f^{-1} in terms of components as $f^{-1} =: (h, g)$

STEP 3: Define W and G as follows:

$$W =: \{x \in \mathbb{R}^n \mid (x, 0) \in V\}$$

(Think of it kind of like an x-axis of V)

$$G(x) =: g(x,0) \text{ for } x \in W$$

Notice W is nonempty since $x_0 \in W$ and W is open since W is just a projection of V on $\mathbb{R}^n$.

Since f^{-1} is differentiable at $(x_0, 0)$ and g is a component of f^{-1} it follows that g is differentiable at $(x_0, 0) \in V$ and so G is differentiable at x_0.

STEP 4: Let's show

$$\{(x,y) \in U \mid F(x,y) = 0\} = \{(x, G(x)) \mid x \in W\}$$

Let A be the left hand side and B be the right-hand-side, and show each set is contained in the other.

$A \subseteq B$: If $(x,y) \in A$ then $(x,y) \in U$ and $F(x,y) = 0$ from which it follows that $f(x,y) = (x, \underbrace{F(x,y)}_{0}) = (x,0)$

Since $(x,0) \in V$ (range of f), by definition $x \in W$ and from $f(x,y) = (x,0)$ we get $(x,y) = f^{-1}(x,0) = (h(x,0), g(x,0))$

Comparing components, this implies $y = g(x,0) = G(x)$ which implies that $(x,y) = (x, G(x))$ and since we've shown $x \in W$, we get that $(x,y) \in B$

Hence $A \subseteq B$ and similarly we have $B \subseteq A$

STEP 5: The only thing left to show is the formula for the derivatives

Note: $G(x_0) = y_0$ because $f(x_0, y_0) = (x_0, 0)$ implies $f^{-1}(x_0, 0) = (x_0, y_0)$ and comparing the second component we get $g(x_0, 0) = y_0$ so $G(x_0) = y_0$

Since $F(x, G(x)) = 0$ for all $x \in W$, F is differentiable at $(x_0, G(x_0)) = (x_0, y_0)$, and G is differentiable at x_0, by the Chain Rule, we have

$$\begin{aligned}
(F(x, G(x)))' &= 0 \\
F_x(x_0, y_0) + F_y(x_0, y_0) G'(x_0) &= 0 \\
F_y(x_0, y_0) G'(x_0) &= -F_x(x_0, y_0) \\
G'(x_0) &= -\left(F_y(x_0, y_0)\right)^{-1} F_x(x_0, y_0) \quad \square
\end{aligned}$$

Note: It is also true that if F is C^k then G is C^k as well.

As mentioned above, the Implicit Function Theorem implies the Inverse Function Theorem, so both of them are equivalent.

3. Application to ODE

Another application of the Banach fixed point theorem is to prove the existence-uniqueness theorem for ODE

In this setting, consider the ODE

$$\begin{cases} \dfrac{du}{dt} = f(u(t)) \\ u(0) = u_0 \end{cases}$$

Picard-Lindelöf Theorem:

If f is Lipschitz and $u_0 \in \mathbb{R}$, then for some small $\tau > 0$, there exists a solution $y : [-\tau, \tau] \to \mathbb{R}$ of the ODE above

Note: The solution is "locally unique," in the sense below.

Compare this to the Cauchy-Peano Theorem. That theorem only required continuity, but gave no uniqueness. This one assumes more, but gives a better result

Proof:[2]

STEP 1: **Main Observation:** By integrating the ODE, it is equivalent to

$$\begin{aligned}\int_0^t u'(s)ds &= \int_0^t f(u(s))ds \\ u(t) - u_0 &= \int_0^t f(u(s))ds \\ u(t) &= u_0 + \int_0^t f(u(s))ds\end{aligned}$$

STEP 2: Let $\tau > 0$ TBA

Since f is continuous, it is bounded around u_0: There is some $r > 0$ and $C > 0$ such that $|f(x)| \leq C$ for all $x \in [u_0 - r, u_0 + r]$.

[2] The proof is a simplified version of the one in Theorem 24 of Pugh's book

Let X be the space of continuous functions $u : [-\tau, \tau] \to [u_0 - r, u_0 + r]$ with the sup norm.

Given $u \in X$, define $\Phi(u) \in X$ (to be shown) by

$$\Phi(u)(t) = u_0 + \int_0^t f(u(s))ds$$

We're done once we show that Φ has a fixed point u, because then $\Phi(u) = u$ and we get

$$u(t) = u_0 + \int_0^t f(u(s))ds \checkmark$$

STEP 3: Proof that Φ is a contraction

First show that $\Phi : X \to X$: Notice that if u is continuous, then $\int_0^t f(u)$ is continuous (in fact differentiable) and hence $\Phi(y)(t)$ is continuous. Moreover

$$|\Phi(u)(t) - u_0| = \left| \int_0^t f(u(s))ds \right| \leq \int_0^t |f(u)| \, ds \leq \int_0^t C ds = Ct \leq C\tau \leq r$$

Provided you choose τ such that $\tau C \leq r$

Hence $\Phi(u) \in [u_0 - r, u_0 + r]$ and so $\Phi(u) \in X$.

Moreover, Φ is a contraction because

$$\begin{aligned}
d(\Phi(y), \Phi(z)) &= \sup_t \left| u_0 + \int_0^t f(y(s))ds - \left(u_0 + \int_0^t f(z(s))ds \right) \right| \\
&\leq \sup_t \left| \int_0^t f(y(s)) - f(z(s))ds \right| \\
&\leq \sup_t \int_0^t |f(y(s)) - f(z(s))| \, ds \\
&\leq \int_0^\tau |f(y(s)) - f(z(s))| \, ds \text{ (the integral is increasing in } t) \\
&\leq \int_0^\tau \left(\sup_s |f(y(s)) - f(z(s))| \right) ds \\
&= \left(\sup_s |f(y(s)) - f(z(s))| \right) \int_0^\tau 1 \\
&\leq L \sup_s |y(s) - z(s)| \, \tau \\
&= L\tau d(y, z)
\end{aligned}$$

This becomes a contraction provided we choose τ so that $L\tau < 1$

STEP 4: Uniqueness

Any other solution $z(t)$ is also a fixed point of Φ, that is $\Phi(z) = z$. Since a contraction has a unique fixed point, we have $z = y$. This is what local uniqueness means. □

4. Uniform Contraction Mapping Principle

The following is a generalization of the Banach fixed point theorem.

Main Idea: Suppose we have family of contraction maps $F(x; \mu)$ indexed by a parameter μ. For each value of the parameter μ, $F(\cdot; \mu)$

has a unique fixed point (by the Banach fixed point theorem). Let $G(\mu)$ map each value of μ to that unique fixed point. The Uniform Contraction Mapping Principle says that the map G is as smooth as the original map F.

We will first need one technical result involving bounds on derivatives of Lipschitz functions.

Fact:

Let X and Y be Banach spaces, $U \subseteq X$ open and $F : U \to Y$ differentiable. If F is Lipschitz with constant L, then

$$\|DF(x)\| \leq L \text{ for all } x \in U$$

Proof: Let $\epsilon > 0$ and $x \in U$. Since F is differentiable at x, we can find $\delta > 0$ such that, whenever $\|h\|_X < \delta$

$$\frac{\|F(x+h) - F(x) - DF(x)h\|_Y}{\|h\|_X} < \epsilon.$$

Let $u \in U$ be a unit vector, and let $h = \frac{\delta}{2} u$ so that $\|h\|_X < \delta$.

Then since $u = h/\|h\|_X$, we have

$$\begin{aligned}
\|DF(x)u\| &= \frac{\|DF(x)h\|_Y}{\|h\|_X} \\
&\leq \frac{\|DF(x)h - (F(x+h) - F(x))\|_Y + \|F(x+h) - F(x)\|_Y}{\|h\|_X} \\
&= \frac{\|F(x+h) - F(x) - DF(x)h\|_Y}{\|h\|_X} + \frac{\|F(x+h) - F(x)\|_Y}{\|h\|_X} \\
&\overset{\text{LIP}}{\leq} \epsilon + \frac{L\|h\|_X}{\|h\|_X} \\
&= \epsilon + L,
\end{aligned}$$

Since ϵ is arbitrary, we have $\|DF(x)\| \leq L$ for all $x \in U$ □

Definition:

$F : D \times B \to D$ is a **uniform contraction** if there is $L < 1$ such that for all $\mu \in B$ and all $u, v \in D$, we have

$$|F(u, \mu) - F(v, \mu)| \leq L|u - v|$$

Definition:

F is **uniformly Lipschitz** in μ if there exists a constant $M > 0$ such that for all $u \in D$ and $\mu_1, \mu_2 \in B$,

$$|F(u, \mu_1) - F(u, \mu_2)| \leq M|\mu_1 - \mu_2|$$

Uniform Contraction Mapping Principle:

Let X be a Banach space and $D \subseteq X$ be a closed and nonempty subset of X

Let Y be a Banach space (parameter space) and $B \subseteq Y$ an open subset of Y

Suppose $F : D \times B \to D$ is a uniform contraction

Let $G : B \to D$ be be the map which associates every $\mu \in B$ with the unique fixed point of $F(\cdot; \mu)$, then:

(1) If F is **uniformly Lipschitz** in μ then G is Lipschitz continuous

(2) If $F \in C^k(D \times B, X)$ for $k \geq 0$, then $G \in C^k(B, X)$

Proof:

STEP 1: Define $G(\mu)$ as in the statement of the theorem. By the Banach fixed point theorem, $G : D \to B$ is the unique function such that $F(x;\mu) = x$ if and only if $x = G(\mu)$.

$$\begin{aligned}|G(\mu_1) - G(\mu_2)| &= |F(G(\mu_1);\mu_1) - F(G(\mu_2);\mu_2)| \\ &\leq |F(G(\mu_1);\mu_1) - F(G(\mu_1);\mu_2)| + |F(G(\mu_1);\mu_2) - F(G(\mu_2);\mu_2)| \\ &\leq |F(G(\mu_1);\mu_1) - F(G(\mu_1);\mu_2)| + L|G(\mu_1) - G(\mu_2)|.\end{aligned}$$

$$(1 - L)|G(\mu_1) - G(\mu_2)| \leq |F(G(\mu_1);\mu_1) - F(G(\mu_1);\mu_2)|.$$

Finally, divide by $(1 - L)$ to get

$$|G(\mu_1) - G(\mu_2)| \leq \frac{1}{1 - L}|F(G(\mu_1);\mu_1) - F(G(\mu_1);\mu_2)|.$$

If F is continuous in both variables (which is part (ii) with $k = 0$, then the RHS above $\to 0$ as $\mu_2 \to \mu_1$, thus G is continuous.

STEP 2: For part (1), if F is uniformly Lipschitz in μ, then the RHS in the previous step becomes

$$|G(\mu_1) - G(\mu_2)| \leq \frac{M}{1 - L}|\mu_1 - \mu_2|$$

STEP 3: All that remains is consider part (2) with $k > 0$. We first consider the case $k = 1$, i.e. F is continuously differentiable.

We will first find a candidate for $DG(\mu)$ and then show that this is the derivative of G

Recall that $G(\mu) = F(G(\mu), \mu)$. If G were differentiable, then we would have

$$DG(\mu) = DF(G(\mu), \mu) = D_X F(G(\mu), \mu) DG(\mu) + D_B F(G(\mu), \mu)$$

This means that $DG(\mu)$ would be a fixed point of the mapping $\Phi : \mathcal{L}(B, X) \times B \to \mathcal{L}(B, X)$, defined by

$$\Phi(A; \mu) = D_X F(G(\mu), \mu) A + D_B F(G(\mu), \mu).$$

The map Φ is a uniform contraction, since for $A_1, A_2 \in \mathcal{L}(B, X)$,

$$\begin{aligned}
&|\Phi(A_1; \mu) - \Phi(A_2; \mu)| \\
=&|D_X F(G(\mu), \mu) A_1 + D_B F(G(\mu), \mu) - (D_X F(G(\mu), \mu) A_2 + D_B F(G(\mu), \mu))| \\
=&|D_X F(G(\mu), \mu)(A_1 - A_2)| \\
\leq&\|D_X F(G(\mu), \mu)\| \, |A_1 - A_2| \\
\leq&L|A_1 - A_2|
\end{aligned}$$

where the last line follows from the previous proposition and the fact that $F(\cdot; \mu)$ is Lipschitz with constant L.

STEP 4: Since $L < 1$, by the Banach fixed point theorem, there exists a function $Z : B \to \mathcal{L}(B, X)$ which maps each $\mu \in B$ to the unique fixed point $Z(\mu)$ of $\Phi(\cdot; \mu)$. Since F is C^1, Φ is continuous, thus by the $k = 0$ case of part (2) of the Uniform Contraction Mapping Principle the map $Z(\mu)$ is continuous.

The function $Z(\mu)$ is our candidate for $DG(\mu)$. All that remains is to use the definition of the derivative to show that $Z(\mu)$ is actually the derivative. This is technical and will be omitted [3]. We then repeat this argument for $k > 1$ for higher order derivatives. □

[3] You can find the complete proof in Lemma 7.2.9 on pages 284-285 of Humpherys, Jarvis, and Evans (2017)