INVERSE FUNCTION THEOREM

FREDERICK TSZ-HO FONG

Theorem 1 (Inverse Function Theorem)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a C^1 function such that at some $a \in \mathbb{R}^n$ we have $\det DF(a) \neq 0$, then there exists an open set U containing a such that $\widetilde{F} := F\big|_U : U \to F(U)$ is bijective with a C^1 inverse $\widetilde{F}^{-1}$.*

Definition 2 (norm of a matrix)**.** Let A be an $n \times n$ matrix of real numbers with (i, j)-th entry a_{ij}. Then we define the L_2-norm of A by:

$$\|A\| := \sqrt{\sum_{1 \le i,j \le n} a_{ij}^2}.$$

Exercise 3. Show that the vector space of $n \times n$ real matrices, denoted by $M_{n\times n}(\mathbb{R})$ equipped with the above $\|\cdot\|$ is a normed vector space.

Lemma 4. *For any $n \times n$ real matrix A and vector $x \in \mathbb{R}^n$, we have*

$$\|Ax\| \le \|A\| \, \|x\| .$$

Here $\|Ax\|$ and $\|x\|$ are the L_2-norm for vectors in $\mathbb{R}^n$, and $\|A\|$ is the L_2-norm of matrices.

Proof. Let a_{ij} be the (i, j)-th entry of A, $x = (x_1, \cdots , x_n)$, and $\{e_i\}_{i=1}^n$ be the standard basis of $\mathbb{R}^n$. Then, we have

$$Ax = \sum_{i=1}^n \left(\sum_{j=1}^n a_{ij} x_j \right) e_i,$$

hence

$$\begin{aligned}
\|Ax\| &= \sqrt{\sum_{i=1}^n \left(\sum_{j=1}^n a_{ij} x_j \right)^2} \\
&\le \sqrt{\sum_{i=1}^n \left(\sum_{j=1}^n a_{ij}^2 \right) \left(\sum_{j=1}^n x_j^2 \right)} \\
&= \sqrt{\sum_{i,j=1}^n a_{ij}^2} \sqrt{\sum_{j=1}^n x_j^2} = \|A\| \, \|x\| .
\end{aligned}$$

□

Lemma 5. *Let $G : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable function on a convex open set Ω in $\mathbb{R}^n$. Then, for any $x, y \in \Omega$, we have*

$$\|G(x) - G(y)\| \le \sqrt{mn} \sup_{z \in \Omega} \|DG(z)\|_\infty \cdot \|x - y\| ,$$

where $\|DG(z)\|_\infty := \max_{i,j} \left| \frac{\partial g_j}{\partial x_i}(z) \right|$

Proof. Consider the straight path $\gamma(t) = (1 - t)x + ty$ which connects x and y. The path is contained in Ω by convexity. Write $G(x) = (g_1(x), \cdots , g_m(x))$, $x = (x_1, \cdots , x_n)$, and $y = (y_1, \cdots , y_n)$, then by the single-variable mean-value theorem, we have for each j:

$$|g_j(x) - g_j(y)| = |g_j \circ \gamma(0) - g_j \circ \gamma(1)| = \left| \frac{d(g_j \circ \gamma)}{dt}(s) \right| |1 - 0|$$

for some $s \in (0,1)$. The chain rule applied to $g_j \circ \gamma$ shows

$$\frac{d(g_j \circ \gamma)}{dt} = \sum_{i=1}^{n} \frac{\partial g_j}{\partial x_i} \frac{d((1-t)x_i + ty_i)}{dt} = \sum_{i=1}^{n} \frac{\partial g_j}{\partial x_i} \cdot (y_i - x_i).$$

By Cauchy-Schwarz, we have

$$\left| \frac{d(g_j \circ \gamma)}{dt}(s) \right| \leq \sqrt{\sum_{i=1}^{n} \left| \frac{\partial g_j}{\partial x_i}(\gamma(s)) \right|^2} \sqrt{\sum_{i=1}^{n} |y_i - x_i|^2} \leq \sqrt{n} \, \|DG(\gamma(s))\|_\infty \, \|x - y\| \, .$$

This proves for each j, we have

$$|g_j(x) - g_j(y)| \leq \sqrt{n} \, \|DG(\gamma(s))\|_\infty \, \|x - y\| \leq \sqrt{n} \sup_{z \in \Omega} \|DG(z)\|_\infty \, \|x - y\| \, ,$$

which implies

$$\|G(x) - G(y)\| = \sqrt{\sum_{j=1}^{m} |g_j(x) - g_j(y)|^2} \leq \sqrt{mn} \sup_{z \in \Omega} \|DG(z)\|_\infty \, \|x - y\| \, .$$

□

Proof of Inverse Function Theorem. Given that $DF(a)$ is invertible, we will pick ε, δ sufficiently small such that any $y \in B_\varepsilon(F(a))$ has a unique $x \in B_\delta(a)$ so that $F(x) = y$. In order to show this, we consider for each $y \in B_\varepsilon(F(a))$ the following map:

$$T_y(x) := x - DF(a)^{-1}(y - F(x)).$$

We will show that by choosing ε and δ sufficiently small, such a map T_y is a contraction map from $\overline{B_\delta(a)}$ to itself. Then by Banach's Contraction Mapping Theorem, such T_y has a unique fixed point $\bar{x}(y) \in B_\delta(a)$, i.e. $T_y(\bar{x}(y)) = \bar{x}(y)$. One can easily check that it implies $y = F(\bar{x}(y))$ as desired.

To verify that T_y is a contraction, we consider

$$\begin{aligned} \|T_y(x_1) - T_y(x_2)\| &= \left\| x_1 - DF(a)^{-1}F(x_1) - x_2 + DF(a)^{-1}F(x_2) \right\| \\ &= \left\| DF(a)^{-1} \left(DF(a)x_1 - F(x_1) \right) - \left(DF(a)x_2 - F(x_2) \right) \right\| \\ &\leq \left\| DF(a)^{-1} \right\| \|G(x_1) - G(x_2)\| \end{aligned}$$

where $G(x) := DF(a)x - F(x)$. Note that $DG(x) = DF(a)I - DF(x) = DF(a) - DF(x)$. Suppose $x_1, x_2 \in B_\delta(a)$, then by Lemma 5, we have

$$\|G(x_1) - G(x_2)\| \leq n \sup_{z \in B_\delta(a)} \|DG(z)\|_\infty \, \|x_1 - x_2\| = n \sup_{z \in B_\delta(a)} \|DF(a) - DF(z)\|_\infty \, \|x_1 - x_2\| \, .$$

Since F is C^1, each entry of $DF(z)$ approaches to the corresponding entry of $DF(a)$ as $z \to a$. Hence, one can choose $\delta > 0$ small so that

$$\|DF(a) - DF(z)\|_\infty \leq \frac{1}{2n \, \|DF(a)^{-1}\|}$$

whenever $z \in B_\delta(a)$, and consequently, we have

$$\|T_y(x_1) - T_y(x_2)\| \leq \left\| DF(a)^{-1} \right\| \|G(x_1) - G(x_2)\| \leq \frac{1}{2} \|x_1 - x_2\| \, . \tag{0.1}$$

This proves $T_y : B_\delta(a) \to \mathbb{R}^n$ is a contraction map.

In order to use the Banach's Contraction Mapping Theorem, we also need to show T_y maps $\overline{B_{\delta/2}(a)}$ to itself. For that we need y being sufficiently close to $F(a)$: first consider

$$\begin{aligned} \|T_y(x) - a\| &\leq \|T_y(x) - T_y(a)\| + \|T_y(a) - a\| \\ &\leq \frac{1}{2} \|x - a\| + \left\| DF(a)^{-1}(y - F(a)) \right\| \\ &\leq \frac{\delta}{4} + \left\| DF(a)^{-1} \right\| \|y - F(a)\| \, . \end{aligned}$$

Let $\varepsilon < \frac{\delta}{4\|DF(a)^{-1}\|}$, then when $y \in B_\varepsilon(F(a))$, we have

$$\|T_y(x) - a\| \leq \frac{\delta}{4} + \frac{\delta}{4} = \frac{\delta}{2}.$$

In other words, we have $T_y(x) \in \overline{B_{\delta/2}(a)}$.

Now by the Banach's Contraction Mapping Theorem applied to $T_y : \overline{B_{\delta/2}(a)} \to \overline{B_{\delta/2}(a)}$ where $y \in B_\varepsilon(F(a))$, there exists a unique fixed point $\bar{x} \in \overline{B_{\delta/2}(a)}$ such that $T_y(\bar{x}) = \bar{x}$, or equivalently, $y = F(\bar{x})$. To summarize, we have proved that by taking $U = F^{-1}(B_\varepsilon(F(a))) \cap B_{\delta/2}(a)$, then

$$\widetilde{F} := F\big|_U : U \subset B_{\delta/2}(a) \to F(U) \subset B_\varepsilon(F(a))$$

is bijective, and one can define an inverse $\widetilde{F}^{-1} : F(U) \to U$.

Next we show that $\widetilde{F}^{-1}$ is continuous on U. By (0.1), we know for any $x_1, x_2 \in U$, we have

$$\begin{aligned}\frac{1}{2}\|x_1 - x_2\| &\geq \left\|(x_1 - DF(a)^{-1}(y - F(x_1))) - (x_2 - DF(a)^{-1}(y - F(x_2)))\right\| \\ &= \left\|(x_1 - x_2) + DF(a)^{-1}(F(x_1) - F(x_2))\right\| \\ &\geq \|x_1 - x_2\| - \left\|DF(a)^{-1}\right\| \|F(x_1) - F(x_2)\|.\end{aligned}$$

This shows

$$\|x_1 - x_2\| \leq 2\left\|DF(a)^{-1}\right\| \|F(x_1) - F(x_2)\|$$

for any $x_1, x_2 \in u$. By writing $x_i = \widetilde{F}^{-1}(y_i)$, we have

$$\left\|\widetilde{F}^{-1}(y_1) - \widetilde{F}^{-1}(y_2)\right\| \leq 2\left\|DF(a)^{-1}\right\| \|y_1 - y_2\|$$

for any $y_1, y_2 \in F(U)$. In other words, $\widetilde{F}^{-1}$ is continuous on $F(U)$.

Next we verify that $\widetilde{F}^{-1}$ is differentiable with Jacobian matrix given by $(DF)^{-1}$. Consider $y_0 = F(x_0) \in F(U)$, and by bijectivity of $\widetilde{F}$ we have write every $y \in F(U)$ as $F(x)$. Then one can check that as $y \to y_0$ (by continuity we have $x \to x_0$ too), we have:

$$\begin{aligned}&\left\|\widetilde{F}^{-1}(y) - \widetilde{F}^{-1}(y_0) - DF(x_0)^{-1}(y - y_0)\right\| \\ &= \left\|\widetilde{F}^{-1}(F(x)) - \widetilde{F}^{-1}(F(x_0)) - DF(x_0)^{-1}(F(x) - F(x_0))\right\| \\ &\leq \left\|x - x_0 - DF(x_0)^{-1}(DF(x_0)(x - x_0) + o\big(\|x - x_0\|\big)\right\| \\ &= \left\|DF(x_0)^{-1}o\big(\|x - x_0\|\big)\right\| \leq o\big(\|x - x_0\|\big).\end{aligned}$$

Hence $\widetilde{F}^{-1}$ is differentiable at any $y_0 \in F(U)$ where $D(\widetilde{F}^{-1})(y_0) = (DF(x_0))^{-1}$.

The fact that $\widetilde{F}^{-1}$ is C^1 follows directly from the fact that its partial derivatives are entries of $(DF(x_0))^{-1}$. By Crammer's rule, each entry of $(DF(x_0))^{-1}$ is a rational function of partial derivatives of F with $\det(DF(x_0))$ as the denominator. Since F is C^1, each entry of $(DF(x_0))^{-1}$ is C^1 too. It completes the proof. $\square$

Remark 6. It is easy to see by induction that if F is C^k, then its local inverse $\widetilde{F}^{-1}$ is also C^k.