Inverse Function Theorem and Implicit Function Theorem

Notation.

1. I_n = $n \times n$ identity matrix.
2. For $\mathbf{x} \in \mathbb{R}^n$, $|\mathbf{x}| = \max\{|x_1|, |x_2|, \ldots, |x_n|\}$ and $\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots x_n^2}$.
3. For a $n \times n$ matrix A , $|A| = \max\{|A_{i,j}|\}$.
4. $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ 이라 하자. f 미분 가능하며 각각의 i, j에 대해 $D_j f_i$가 연속이면 $f \in C^1$ **class** function 이라 한다. For each i, j, and $\mathrm{x} \in A$, $D_j f_i(\mathrm{x})$ 가 $r-1$번 미분 가능하며 그 $r-1$번째 도함수가 연속이면 f를 C^r **class** function 이라 한다.

Lemma 1.

$n \times m$ 행렬 A와 $m \times p$ 행렬 B에 대해 $|A \cdot B| \le m|A||B|$ 이다.

Lemma 2.

A 가 open in $\mathbb{R}^n$, B 가 open in $\mathbb{R}^m$ 이며 $f : A \to \mathbb{R}^m$, $g : B \to \mathbb{R}^p$, and $f(A) \subset B$라 하자. f, g가 C^r class function 이면 그 합성함수 $g \circ f$도 C^r class function 이다.

(Proof) Induction을 통해 증명한다. $f, g \in C^1$ 이라 하자. 우리는 $D(g \circ f)(\mathrm{x}) = Dg(f(\mathrm{x})) \cdot Df(\mathrm{x})$ 임을 알고 있다. $g \in C^1$ 이므로 Dg 는 연속함수이다. f가 연속함수 이므로 $Dg(f)$는 연속함수이다. $f \in C^1$ 이므로 $D(g \circ f)$ 는 연속함수이다. 따라서 $D(g \circ f) \in C^1$ 이다.

이제 $f, g \in C^{r-1}$ 이면 $g \circ f \in C^{r-1}$ 이라 가정하자. 즉 $D_j g_i(f(\mathbf{x})) \in C^{r-1}$ 이다. $f, g \in C^r$ 이라 하자. $Dg(\mathrm{y})$ 와 $Df(\mathbf{x})$ 가 $r-1$ 번 미분 가능 하며 따라서 $D(g \circ f)(\mathbf{x}) = Dg(f(\mathbf{x})) \cdot Df(\mathbf{x})$ 도 $r-1$ 번 미분 가능 하므로 $g \circ f$도 C^r class function 이다.

Theorem 3. (Mean value theorem for $\mathbb{R}^n$)

Let A be open in $\mathbb{R}^n$ and $f : A \to \mathbb{R}$ be differentiable on A. 만약 $\mathbf{a}$ 에서 $\mathbf{a} + \mathbf{h}$ 로의 line segment가 A에 포함된다면 어떤 $\mathbf{c} = \mathbf{a} + t_0\mathbf{h}$ with $0 < t_0 < 1$ 에서 $f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = Df(\mathbf{c}) \cdot \mathbf{h}$ 이다.

Proof is trivial

Comment : $\mathbf{a} \in A$ 를 중심으로 A에 포함되는 open cube나 open ball 형태의 neighborhood N_a를 잡으면 모든 $\mathbf{x}, \mathbf{y} \in N_a$를 잇는 line segment가 N_a의 subset이므로 mean value theorem이 성립한다.

Lemma 4.

A is open in $\mathbb{R}^n$, $f : A \to \mathbb{R}^n$ is a C^1 class function 이라 하자. $Df(\mathbf{a})$ 가 non singular 이면 $^\exists\alpha > 0$ s.t $|f(\mathbf{x}_0) - f(\mathbf{x}_1)| \geq \alpha|\mathbf{x}_0 - \mathbf{x}_1|$ holds for all $\mathbf{x}_0,\, \mathbf{x}_1$ in some open cube $C(\mathbf{a},\, \varepsilon)$ centered at $\mathbf{a}$. 즉 $f(\mathbf{x})$ is injective in $C(\mathbf{a},\, \varepsilon)$.

(Proof) $E = Df(\mathbf{a})$ 라 하자. $|\mathbf{x}_0 - \mathbf{x}_1| = |E^{-1} \cdot E \cdot (\mathbf{x}_0 - \mathbf{x}_1)| \leq n|E^{-1}| \cdot |E \cdot (\mathbf{x}_0 - \mathbf{x}_1)|$.

Let $2\alpha = 1/(n|E^{-1}|)$, then $|E \cdot (\mathbf{x}_0 - \mathbf{x}_1)| \geq 2\alpha|\mathbf{x}_0 - \mathbf{x}_1|$.

Consider function $H(\mathbf{x}) = f(\mathbf{x}) - E \cdot \mathbf{x}$, then $DH(\mathbf{a}) = 0$. H가 C^1 함수이므로 $DH(\mathbf{x}) < \alpha/n$ for all $x \in C(\mathbf{a},\, \varepsilon) = C$ 이 되도록 하는 $\varepsilon > 0$이 존재한다. Mean value theorem에 의해 $\mathbf{x}_0,\, \mathbf{x}_1 \in C$ 이면 $^\exists\mathbf{c} \in C$ such that $|H_i(\mathbf{x}_0) - H_i(\mathbf{x}_1)| = |DH_i(\mathbf{c}) \cdot (\mathbf{x}_0 - \mathbf{x}_1)| \leq n(\alpha/n)|\mathbf{x}_0 - \mathbf{x}_1|$.

따라서, 모든 $\mathbf{x}_0,\, \mathbf{x}_1 \in C$ 에 대해

$$\begin{aligned}\alpha|\mathbf{x}_0 - \mathbf{x}_1| &\geq |H(\mathbf{x}_0) - H(\mathbf{x}_1)| \\ &= |f(\mathbf{x}_0) - E \cdot \mathbf{x}_0 - f(\mathbf{x}_1) + E \cdot \mathbf{x}_1| \\ &\geq |E \cdot \mathbf{x}_1 - E \cdot \mathbf{x}_0| - |f(\mathbf{x}_1) - f(\mathbf{x}_0)| \\ &\geq 2\alpha|\mathbf{x}_1 - \mathbf{x}_0| - |f(\mathbf{x}_1) - f(\mathbf{x}_0)|\ .\end{aligned}$$

이므로 Lemma가 성립한다. □.

Lemma 5.

Let A be an open in $\mathbb{R}^n$ and $\phi : A \to \mathbb{R}$ be differentiable. 만약 ϕ가 $\mathbf{x}_0 \in A$에서 local minimum을 가지면 $D\phi(\mathbf{x}_0) = 0$ 이다.

Proof is trivial

Theorem 6.

Let A be open in $\mathbb{R}^n$, $f : A \to \mathbb{R}^n$ be of class C^r, and $B = f(A)$. 만약 f가 one-to-one on A 이고 $Df(\mathbf{x})$ 가 non singular for $\mathbf{x} \in A$ 이면 B는 open in $\mathbb{R}^n$ 이며 $f^{-1} = g$ 는 C^r class 함수이다.

(Proof) (Step 1) 우선 B가 open 임을 보이자. 임의의 $\mathbf{b} \in B$ 에 대해 open ball $B(\mathbf{b},\, \delta) \subset B$가 존재함을 보이고자 한다. $\mathbf{a} = f^{-1}(\mathbf{b})$ 를 내부에 포함하며 A에 포함되는 closed rectangle Q를 생각하자. ($\mathbf{a} \in \mathrm{int}(Q)$ and $Q \subset A$). $\mathrm{Bd}(Q)$는 compact set 이고 f는 연속이므로 $f(\mathrm{Bd}(Q))$도 compact set 이다. f가 injection 이므로 $\mathbf{b} \notin \mathrm{Bd}(Q)$ 이다. $f(\mathrm{Bd}(Q))$가 closed set 이므로 이것과 disjoint 한 open ball around $\mathbf{b}$, $B(\mathbf{b},\, 2\delta)$ 가 존재한다. 이제 임의의 $\mathbf{c} \in B(\mathbf{b},\, \delta)$에 대해 $\mathbf{c} = f(\mathbf{x})$ for some $\mathbf{x} \in A$ 임을 보이자.

이를 위해 C^r class 함수 $\phi(\mathbf{x}) = \|f(\mathbf{x}) - \mathbf{c}\|^2$ 를 생각하자. Q가 compact 하므로 ϕ는 Q에서 최대값과 최소값을 가진다. $\mathbf{c} \in B(\mathbf{b},\, \delta)$ 이므로 $\phi(\mathbf{a}) = \|f(\mathbf{a}) - \mathbf{c}\|^2 = \|\mathbf{b} - \mathbf{c}\|^2 < \delta^2$ 이다. 따라서 minimum value of ϕ on Q 는 δ^2 보다 작아야 한다. 그런데 $\mathbf{y} \in \mathrm{Bd}(Q)$ 이면 $f(\mathbf{y})$ 는 $B(\mathbf{b},\, 2\delta)$ 밖에 있으므로 $\phi(\mathbf{y}) \geq \delta^2$. 따라서 ϕ를 minimum 이 되도록 하는 값 $\mathbf{x}$ 는 $\mathrm{int}(Q)$에 존재한다.

f 는 $\mathbf{x} \in \mathrm{int}(Q)$ 에서 local minimum을 가지므로 $D\phi(\mathbf{x}) = 0$. (Lemma 4.) $D_j\phi(\mathbf{x}) = 0$ for all j.
$D_j\phi(\mathbf{x}) = \sum_{k=1}^{n} 2\left(f_k(\mathbf{x}) - c_k\right) D_j f_k(\mathbf{x})$ 이므로 $D\phi(\mathbf{x}) = 0$ 을 행렬방정식으로 쓰면

$$2Df(\mathbf{x}) \cdot \begin{bmatrix} f_1(\mathbf{x}) - c_1 \\ \vdots \\ (f_n\mathbf{x}) - c_n \end{bmatrix} = 0$$

이 된다. $Df(\mathbf{x})$가 non-singular 이므로 $f(\mathbf{x}) - \mathbf{c} = 0$인 x가 $B(\mathbf{b},\, \delta)$에 존재한다.

(Step 2) 이제 $g = f^{-1}$이 연속임을 보이자. g가 연속인 것은 임의의 open $U \subset A$에 대해 $f(U)$가 open in B 임을 보이면 되는데 이는 step 1에서 보인것이다. 따라서 $g = f^{-1}$는 연속이다.

(Step 3) g 가 임의의 $\mathbf{b} \in B$에서 differentiable 임을 보이자. $\mathbf{a} = g(\mathbf{b})$ 이고 $E = Df(\mathbf{a})$ 라 하자.
$N_0 = \{\mathbf{x} \in A : 0 < \|x\| < r, \text{for some } r\ \}$ 라 하면 N_0 는 open in A 이다. $G : N_0 \to B$ 를 다음과 같이 정의한다.

$$G(\mathbf{k}) = \frac{\left[g(\mathbf{b}+\mathbf{k}) - g(\mathbf{b}) - E^{-1} \cdot \mathbf{k}\right]}{|\mathbf{k}|}\ .$$

g 는 $\mathbf{b}$ 에서 미분가능하며 $Dg(\mathbf{b}) = E^{-1}$ 이다. $\Delta(\mathbf{k}) = g(\mathbf{b}+\mathbf{k}) - g(\mathbf{b})$ 라 정의한다.

Lemma 4 로부터 $\mathbf{a}$의 cubic neighborhood C와 $\alpha > 0$ such that $|f(\mathbf{x}_0) - f(\mathbf{x}_1) \geq \alpha|\mathbf{x}_0 - \mathbf{x}_1|$ for all $\mathbf{x}_0,\, \mathbf{x}_1 \in C$ 이 존재함을 알고 있다. $f(C)$가 $\mathbf{b}$의 neighborhood 이다. $|\mathbf{k}| < \varepsilon$ 일 때 $f(\mathbf{b}+\mathbf{k}) \in f(C)$ 이도록 ε을 작게 잡자. 그렇다면 $g(\mathbf{b}+\mathbf{k}) = \mathbf{x}_0$, $g(\mathbf{b}) = \mathbf{x}_1$ 가 되도록 할 수 있으며 다음이 성립한다.

$$|(\mathbf{b}+\mathbf{k}) - \mathbf{b}| \geq \alpha|g(\mathbf{b}+\mathbf{k}) - g(\mathbf{b})|\ .$$

따라서 $1/\alpha \geq |\Delta(\mathbf{k})|/|\mathbf{k}|$ 이다. 즉 우리는 $\Delta(\mathbf{k})|/|\mathbf{k}|$ 가 bounded 되도록 하는 $|\mathbf{k}| < \varepsilon$을 생각 할 수 있다.

이제 $\mathbf{k} \to 0$ implies $G(\mathbf{k}) \to 0$ 임을 보이자. Let $0 < \mathbf{k} < \varepsilon$ 이라 하자. g 가 injection 이므로 $\mathbf{k} \neq 0$ 이면 $\Delta(\mathbf{k}) \neq 0$ 임을 이용하면

$$G(\mathbf{k}) = \frac{\Delta(\mathbf{k}) - E^{-1} \cdot \mathbf{x}}{|\mathbf{k}|} = -E \cdot \left[\frac{\mathbf{k} - E \cdot \Delta(\mathbf{k})}{|\Delta(\mathbf{k})|}\right] \cdot \frac{|\Delta(\mathbf{k})|}{|\mathbf{k}|}\ .$$

E^{-1}는 constant, $|\Delta(\mathbf{k})|/|\mathbf{k}|$ 는 bounded 이다. 위 식의 [ ] 부분이 $\mathbf{k} \to 0$ 일 때 0에 접근함을 보이자. $\mathbf{b}+\mathbf{k}$ $= f(g(\mathbf{b}+\mathbf{k})) = f(g(\mathbf{b}) + \Delta(\mathbf{k})) = f(\mathbf{a} + \Delta(\mathbf{k}))$ 이므로, 위 식의 [ ] 부분은 다음과 같다.

$$\frac{f(\mathbf{a} + \Delta(\mathbf{k})) - f(\mathbf{a}) - E \cdot \Delta(\mathbf{k})}{|\Delta(\mathbf{k})|}$$

$\mathbf{k} \to 0$ implies $\Delta(\mathbf{k}) \to 0$ because g is continuous. 따라서 f는 $\mathbf{a}$ 에서 미분가능 하며 그 derivative는 E 이다.

(Step 5) 이제 f가 C^r class function 이면 $g = f^{-1}$ 도 C^r class function 임을 induction을 통해 보인다. f는 C^1 함수이므로 Df는 연속이다. $Dg(\mathbf{y}) = [Df(g(\mathbf{y}))]^{-1}$ 이며 g, Df와 I_n이 연속이므로 $Dg(\mathbf{y})$도 연속이다. 따라서 $g \in C^1$. 이제 f가 C^{r-1} 이면 $g = f^{-1} \in C^{r-1}$ 이라 가정하자. $Df(g(\mathbf{y}))$ 도 C^{r-1} class function 이므로 $g\ \in C^r$ 이다.

Theorem 7 (Inverse Function Theorem)

A 가 open in $\mathbb{R}^n$ 이며 $f : A \to \mathbb{R}^n$ 이 C^r class 함수라 하자. $Df(\mathbf{x})$ 가 non-singular at $\mathbf{a} \in A$ 이면 $\mathbf{a}$ 의 어떤 neighborhood N_a에서 $f|_{N_a} : N_a \to \mathbb{R}^n$은 injection 이며 $g : f(N_a) \to N_a$ defined by $g = (f|_{N_a})^{-1}$ 은 C^r class function 이다.

(Proof) Lemma 4에 의해 f가 injection 인 $\mathbf{a}$의 neighborhood U_1이 존재한다. $Df(\mathbf{a})$ 가 non singular and continuous at $\mathbf{a}$ 이므로 $Df(\mathbf{x})$가 nonsingular 한 $\mathbf{a}$ 의 neighborhood U_2가 존재한다. $N_a = U_1 \cap U_2$라 하면 Theorem 6으로 부터 Inverse function theorem 이 성립함을 알 수 있다. □

Definition.

A가 open in $\mathbb{R}^m$ 이고 $f : A \to \mathbb{R}^n$ 이 differentiable 이며 $f_1, \ldots, f_n$ 이 f 의 component function 이라 하자. 이 때 Df를 다음과 같이 쓰기도 한다.

$$Df = \frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_m)} = \frac{\partial f}{\partial \mathbf{x}}$$

이 때 $(Df)_{i,j} = \partial f_i / \partial x_j$ 이다.

Theorem 8.

A 가 open in $\mathbb{R}^{k+n}$ 이고 $f : A \to \mathbb{R}^n$ 이 differentiable이라 하자. f를 $\mathbf{x} \in \mathbb{R}^k$, $\mathbf{y} \in \mathbb{R}^n$ 에 대해 $f(\mathbf{x}, \mathbf{y})$ 로 쓰기로 하자. 그렇다면 $Df = [\partial f / \partial \mathbf{x}, \partial f / \partial \mathbf{y}]$ 로 쓸 수 있다. Open B in $\mathbb{R}^k$ 에 대해 differentiable function $g : B \to \mathbb{R}^n$ 이 존재하여 $f(\mathbf{x}, g(\mathbf{x})) = 0$ for all $\mathbf{x} \in B$ 라 하자. 그렇다면 $\mathbf{x} \in B$에서 다음이 성립한다.

$$\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}, g(\mathbf{x})) + \frac{\partial f}{\partial \mathbf{y}}(\mathbf{x}, g(\mathbf{x})) \cdot Dg(\mathbf{x}) = 0 \ .$$

(Proof) Define $h : B \to \mathbb{R}^{k+n}$ and $H : B \to \mathbb{R}^n$ by $h(\mathbf{x}) = (\mathbf{x}, g(\mathbf{x}))$, $H(\mathbf{x}) = f(h(\mathbf{x})) = f(\mathbf{x}, g(\mathbf{x})) = 0$.

From chain rule

$$0 = DH(\mathbf{x}) = Df(h(\mathbf{x}, g(\mathbf{x})) = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{x}}(h(\mathbf{x})) & \frac{\partial f}{\partial \mathbf{y}}(h(\mathbf{x})) \end{bmatrix} \cdot \begin{bmatrix} I_k \\ Dg(\mathbf{x}) \end{bmatrix} = \frac{\partial f}{\partial \mathbf{x}}(h(\mathbf{x})) + \frac{\partial f}{\partial \mathbf{y}}(h(\mathbf{x})) \cdot Dg$$

□.

<u>Theorem 9. (Implicit Function Theorem)</u>

A가 open in $\mathbb{R}^{k+n}$ 이고 $f : A \to \mathbb{R}^n$ 이 C^r class function이라 하자. 앞서와 같이 $\mathbf{x} \in \mathbb{R}^k$, $\mathbf{y} \in \mathbb{R}^n$ 에 대해 $f = f(\mathbf{x}, \mathbf{y})$ 로 쓰기로 하자. $(\mathbf{a}, \mathbf{b}) \in A$ 이며 $f(\mathbf{a}, \mathbf{b}) = 0$ 이고 $\det(\partial f / \partial \mathbf{y})(\mathbf{a}, \mathbf{b}) \neq 0$ 이면 $\mathbf{a}$의 어떤 neighborhood $U \subset \mathbb{R}^k$ 에서 C^r 함수 $g : U \to \mathbb{R}^n$ 이 존재하여 $f(\mathbf{x}, g(\mathbf{x})) = 0$ for all $\mathbf{x} \in U$ 이다. 이 g는 unique 하다.

(Proof) Define $F : A \to \mathbb{R}^{k+n}$ by $F(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, f(\mathbf{x}, \mathbf{y}))$. Then,

$$DF = \begin{bmatrix} I_k & 0 \\ \partial f / \partial \mathbf{x} & \partial f / \partial \mathbf{y} \end{bmatrix} \ .$$

여기서 $\det(DF) = \det \partial f / \partial \mathbf{y}$ 이므로 DF는 non-singular at $(\mathbf{a}, \mathbf{b})$ 이다.

$F(\mathbf{a},\, \mathbf{b}) = (\mathbf{a},\, 0)$ 이므로 inverse function theorem 을 사용하면 $(\mathbf{a},\, \mathbf{b})$의 neighborhood $U \times V \subset \mathbb{R}^{k+n}$ 이 존재하여 $F(U \times V)$ 는 $(\mathbf{a},\, 0)$ 의 neighborhood 이며(Let it W) , $F|_{U\times V}$ 는 injection이다. 또한 $(F|_{U\times V})^{-1} = G : W \to U \times V$ 가 존재하며 C^r class function 이다. 따라서 $(\mathbf{x},\, \mathbf{y}) = G(\mathbf{x},\, f(\mathbf{x},\, \mathbf{y}))$ 이다. G는 F처럼 첫번째 k coordinate를 보존하며 따라서 $G(\mathbf{x},\, \mathbf{z}) = (\mathbf{x},\, h(\mathbf{x},\, \mathbf{z}))$ 로 쓸 수 있다. 여기서 $h : W \to \mathbb{R}^n$ 은 C^r 함수이다.

$\mathbf{a}$의 connected neighborhood B를 $B \times \mathbf{0} \in W$ 가 되도록 잡을 수 있다. $\mathbf{x} \in B$ 라면,

$$
\begin{aligned}
G(\mathbf{x},\, \mathbf{0}) &= ((x),\, h(\mathbf{x},\, \mathbf{0}))\ , \\
(\mathbf{x},\, \mathbf{0}) &= F(\mathbf{x},\, h(\mathbf{x}, \mathbf{0})) = (\mathbf{x},\, f(\mathbf{x},\, h(\mathbf{x},\, \mathbf{0})))\ , \\
\mathbf{0} &= f(\mathbf{x},\, h(\mathbf{x},\, \mathbf{0}))\ .
\end{aligned}
$$

$g(\mathbf{x}) = h(\mathbf{x},\, 0)$으로 정의하면 우리는 $f(\mathbf{x},\, g(\mathbf{x})) = 0$이므로 우리가 원하는 함수임을 알 수 있다. 더우기 $(\mathbf{a},\, \mathbf{b}) = G(\mathbf{a},\, \mathbf{0}) = (\mathbf{a},\, h(\mathbf{a},\, \mathbf{0}))$ 이므로 $\mathbf{b} = h(\mathbf{a})$ 이다.

이제 g의 uniqueness를 보이자. $g' : B \to \mathbb{R}^n$ 이 주어진 조건을 만족하는 다른 함수라 하자. $g(\mathbf{a}) = g'(\mathbf{a}) = 0$ 이다. $g(\mathbf{a}) = \mathbf{b} \in V$ 이며 V 는 open, g'은 연속 이므로 $(g')^{-1}(V) = B_0$ 는 $\mathbf{a}$의 neighborhood 이다. $f(\mathbf{x},\, g'(\mathbf{x})) = 0$ for all $\mathbf{x} \in B_0$ 이므로

$$
\begin{aligned}
F(\mathbf{x},\, g'(\mathbf{x})) &= (\mathbf{x},\, 0)\ ,\ \text{so} \\
(\mathbf{x},\, g'(\mathbf{x})) &= G(\mathbf{x},\, 0) = (\mathbf{x},\, h(\mathbf{x})).
\end{aligned}
$$

이므로 $g = g'$ on B_0.