1 Twierdzenie o lokalnej odwracalności

1.1 Wstęp motywacyjny

Dla funkcji jednej zmiennej mieliśmy *twierdzenie o funkcji odwrotnej*, które dla wygody tutaj przypomnimy.

Niech $f : \mathbb{R} \supset]a, b[\to \mathbb{R}$; przyjmijmy, że f jest klasy C^1.

Tw. (1-1) Jeśli $f'(x_0) \neq 0$, to wtedy $f'(x) \neq 0$ na pewnym przedziale $]x_0 - \epsilon, x_0 + \epsilon[$ i wtedy f jest *bijekcją* odcinka $]a, b[$ na $]f(a), f(b)[$ (dla $f'(x_0) > 0$; w przypadku gdy $f'(x_0) < 0$ bijekcja jest na odcinek $]f(b), f(a)[$). Innymi słowy, w każdym punkcie $x \in]x_0 - \epsilon, x_0 + \epsilon[$ istnieje funkcja odwrotna f^{-1}.

Zastanówmy się teraz, czy i jak można to twierdzenie rozszerzyć na przypadek *odwzorowań* $f : \mathbb{R}^n \to \mathbb{R}^m$ klasy C^1.

Przykłady z zakresu odwzorowań liniowych pokazują, kiedy na pewno *nie jest* możliwe znalezienie odwzorowania odwrotnego. I tak, dla odwzorowania $\mathbb{R}^2 \to \mathbb{R}$, pytanie o istnienie odwz. odwrotnego jest równoważne pytaniu o istnienie jednoznacznego rozwiązania układu równań liniowych. I tak np. układ równań $ax + by = A$ *nie ma* jednoznacznego rozwiązania (gdy $b \neq 0$, to rozwiązaniem jest: $(x, y = \frac{A-ax}{b})$ dla dowolnego x; a gdy $b = 0$, to jest rozwiązaniem jest $(x = \frac{A}{a}, y)$-dowolne; w obu więc przypadkach nie ma jednoznaczności rozwiązania). Podobnie można się przekonać, że dla $n < m$ żadne odwzorowanie liniowe nie może być wzajemnie jednoznaczne.

Przyjmijmy więc, że $n = m$, i spróbujmy zahipotezować kryteria na odwracalność odwzorowania.

Rozpatrzmy najsampierw możliwie prostą sytuację odwzorowania $\mathbb{R}^n \to \mathbb{R}^n$. Niedawno mieliśmy do czynienia z odwzorowaniami *liniowymi* $A : \mathbb{R}^n \to \mathbb{R}^n$. Niech $y = Ax$. Wtedy pytanie, czy odwzorowanie A jest odwracalne (tzn. czy istnieje A^{-1}, jest równoważne pytaniu, czy *układ równań liniowych* $Ax = y$ posiada *jednoznaczne* rozwiązanie. Niedawno pokazaliśmy twierdzenia (tw. Kroneckera-Capellego, z doprecyzowaniem danym przez wzory Cramera) że odpowiedź jest pozytywna w przypadku, gdy macierz A jest odwracalna, co jest równoważne temu, że $\det A \neq 0$. To więc będzie nasz drogowskaz.

Analogonem pochodnej funkcji jest pochodna odwzorowania, tzn. macierz Jacobiego. Można przypuścić, że gdy macierz Jacobiego będzie *nieosobliwa* (tzn. rząd macierzy Jacobiego będzie równy n, lub – równoważnie – wyznacznik macierzy Jacobiego będzie różny od zera), to odwzorowanie lokalnie da się odwrócić – zachodzi analogon powyższego Tw. (1-1). Nieco dokładniej, odwracalność zachodzi na *małym otoczeniu* punktu x_0 w przeciwobrazie i *małym otoczeniu* $T(x_0)$ w obrazie. Do takiej *lokalnej* odwracalności wystarczy, aby macierz Jacobiego w punkcie x_0 była nieosobliwa.

Uwaga. Przypadek odwzorowań $\mathbb{R}^n \to \mathbb{R}^n$ jest jednak istotnie inny niż funkcji $\mathbb{R}^1 \to \mathbb{R}^1$, i różne twierdzenia, które zachodziły dla funkcji, nie przenoszą się na odwzorowania. Np. dla funkcji zachodzi tw. będące mocniejszą wersją tw. (1-1). Stało tam, że jeżeli funkcja f jest różniczkowalna i jej pochodna jest wszędzie różna od zera na jakimś odcinku $[a, b]$, to jest bijekcją odcinka $[a, b]$ na $[f(a), f(b)]$ (**RYS.**). Sytuacja w przypadku odwzorowań jest inna.

Przykł. Rozpatrzmy: $\mathbb{R}^2 \ni (x, y) \to (e^x \cos y, e^x \sin y) \in \mathbb{R}^2$. Mamy:

$$T'(x) = \begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix}, \qquad \det T'(x, y) = e^x \neq 0$$

zatem jakobian odwzorowania wszędzie jest różny od zera i macierz Jacobiego jest wszędzie nieosobliwa. Ale: $T(x, y+2\pi) = T(x, y)$, tzn. odwzorowanie *nie jest* globalnie odwracalne. Tzn: Każdy punkt przestrzeni $\mathbb{R}^2$ ma otoczenie, w którym odwzorowanie T jest odwracalne. Pomimo tego nie jest ono wzajemnie jednoznaczne na całym $\mathbb{R}^2$.

Okazuje się, że jest niełatwo podać warunki na *globalną* odwracalność odwzorowań. Na razie będziemy się jednak zadowalać się odwracalnością *lokalną*, co wystarczy nam do zastosowań takich, jak zamiana zmiennych w całkach wielokrotnych, lub w równaniach różniczkowych.

Dokładniejsze sformułowanie podamy później, a na razie zdefiniujemy pojęcia, które będą potrzebne dalej przy dowodzie.

1.2 Norma na przestrzeni macierzy

Własności normy na przestrzeni wektorowej. Pamiętamy, że jeżeli $x, y \in \mathbb{R}^N$ – wektory (tzn. $x = (x^1, x^2, \ldots, x^N)$, a λ – liczba, i określiliśmy normę $||x||$ wzorem

$$||x|| = \sqrt{\sum_{k=1}^{N} (x^i)^2}$$

to norma posiada własności:

1. $||\lambda x|| = |\lambda| \cdot ||x||$ dla dowolnego $\alpha \in \mathbb{R}$ i $x \in \mathbb{R}^N$.
2. $||x|| \geqslant 0$, a równość $||x|| = 0$ zachodzi tylko dla wektora zerowego $x = 0$.
3. $||x + y|| \leqslant ||x|| + ||y||$

W przestrzeni $\mathbb{R}^N$ możemy też wprowadzić *iloczyn skalarny*[1]. Iloczynem skalarnym wektorów $x, y \in \mathbb{R}^N$ nazywamy liczbę $(x|y)$ określoną jako

$$(x|y) = \sum_{k=1}^{N} x^i y^i \tag{1}$$

Dla tak zdefiniowanego iloczynu skalarnego zachodzi *nierówność Schwarza*:

$$\left| \sum_{k=1}^{N} x^i y^i \right| \leqslant ||x|| \cdot ||y||.$$

Jak pamiętamy sprzed kilku wykładów z części algebraicznej (a jeśli nie pamiętamy to niniejszym przypominamy), że zbiór macierzy ustalonego rozmiaru jest *przestrzenią wektorową*. Załóżmy, że mamy do czynienia z macierzami rozmiaru $m \times n$, tzn. o m wierszach i n kolumnach. Taką macierz A zapisujemy jako: $A = ({a^i}_j)$, gdzie $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$. Gdy chcemy z macierzy A 'wyjąć' element macierzowy ${a^i}_j$, to zapisujemy to jako: ${(A)^i}_j$ (element na skrzyżowaniu i-tego wiersza i j-tej kolumny). Macierze można *dodawać*: Jeśli A, B – macierze $m \times n$, to ich suma $C = A + B$ jest macierzą $m \times n$ o elementach

$${c^i}_j = {a^i}_j + {b^i}_j;$$

[1] Na razie jest to jedynie definicja i nazwa; iloczyn skalarny ma kilka własności, które będą wymienione później

macierz można także *pomnożyć przez liczbę*: Jeśli A – macierz, λ – liczba, to iloczynem λA nazywamy macierz o elementach

$$(\lambda A)^i{}_j = \lambda a^i{}_j.$$

Te dwie operacje (dodawanie macierzy oraz mnożenie macierzy przez liczbę) czynią ze zbioru macierzy (oznaczamy go $\mathbb{R}^m{}_n$) przestrzeń wektorową. Wymiar tej przestrzeni wynosi tyle, ile jest niezależnych elementów macierzowych; tu wszystkie są niezależne, więc wymiar wynosi: $\dim \mathbb{R}^m{}_n = m \cdot n$.

Przypomnijmy jeszcze, jak macierz A działa na wektor $x \in \mathbb{R}^n$: Wynikiem jest wektor $y = Ax = (y^{\cdot} \ldots, y^m) \in \mathbb{R}^m$ o składowych

$$y^i = (Ax)_i = \sum_{k=1}^{n} a^i{}_j x^j \tag{2}$$

Do zdefiniowania normy na przestrzeni macierzy będzie nam potrzebny następujaćy

Lemat. Niech $A = (a^i{}_j)$ – macierz $m \times n$. Istnieje wtedy taka stała $C \geqslant 0$, że dla dowolnego wektora $x = (x^1, \ldots, x^n) \in \mathbb{R}^n$ zachodzi nierówność

$$\underbrace{||Ax||}_{\text{liczona w } \mathbb{R}^m} \leqslant C \cdot \underbrace{||x||}_{\text{liczona w } \mathbb{R}^m} \tag{3}$$

Dow. Niech $y = Ax$ liczone jak w (2). Liczymy kwadrat normy wektora Ax:

$$||Ax||^2 = ||y||^2 = \sum_{i=1}^{m} (y^i)^2 = \sum_{i=1}^{m} (\sum_{j=1}^{n} a^i{}_j x^j)^2 = \spadesuit$$

Potraktujmy teraz macierz A jako kolekcję m wektorów (wierszowych) o długości n. Tzn. $j-$ta składowa i-tego wektora a_i to $a^i{}_j$. W ten sposób, drugą sumę w powyższej podwójnej sumie można potraktować jako *iloczyn skalarny* wektorów a_i oraz x. Korzystając z nierówności Schwarza w $\mathbb{R}^n$, mamy:

$$\spadesuit \leqslant \sum_{i=1}^{m} ||a_i||^2 \cdot ||x||^2 = ||x||^2 \sum_{i=1}^{m} \left(\sum_{j=1}^{n} (a^i{}_j)^2 \right) = ||x||^2 \sum_{i=1}^{m} \sum_{j=1}^{n} (a^i{}_j)^2.$$

Wyciągając pierwiastek (obie strony są nieujemne), mamy

$$||Ax|| \leqslant \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (a^i{}_j)^2} \cdot ||x||;$$

za liczbę C w sformułowaniu Lematu możemy więc wziąć np.

$$C = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (a^i{}_j)^2}.$$

CBDO

Def. *Normą* macierzy A nazywamy kres dolny zbioru $\Gamma \subset \mathbb{R}_+ \cup \{0\}$, gdzie

$$\Gamma = \{C \in \mathbb{R}_+ \cup \{0\} : \underset{x \in \mathbb{R}^n}{\forall} ||Ax|| \leqslant C||x||\}$$

Uwaga. Definicja jest z sensem, bo z pokazanego dopiero co Lematu wynika, że zbiór Γ jest niepusty. Stąd też mamy *nieujemność* normy:

$$||A|| \geqslant 0$$

dla dowolnej macierzy A.

Stw. Norma macierzy posiada następujące własności:

0. $\underset{x}{\forall} ||Ax|| \leqslant ||A|| \cdot ||x||$.
1. $||\lambda A|| \leqslant |\lambda| \cdot ||A||$.
2. $||A + B|| \leqslant ||A|| + ||B||$.
3. $||A|| \geqslant 0$, przy czym $||A|| = 0 \iff A = 0$.
4. $||A \cdot B|| \leqslant ||A|| \cdot ||B||$ dla A, B – takich, że iloczyn $A \cdot B$ jest określony.

Dow.

0. Wynika z definicji normy.
1. Oczywiste.
2. Policzmy $||(A + B)x||$:

   $$||(A+B)x|| = \underbrace{||Ax + Bx||}_{\text{norma wektora}} \overbrace{\leqslant}^{\text{v.S.i.}} ||Ax||+||Bx|| \overbrace{\leqslant}^{\text{wl. 0.}} ||A||\cdot||x||+||B||\cdot||x|| = (||A||+||B||)||x||$$

   Pokazaliśmy więc, że dla dowolnego wektora x zachodzi $||(A+B)x|| \leqslant (||A||+||B||)\cdot ||x||$. Porównajmy to z definicją normy: Norma $(A+B)$ to *kres dolny* liczb C takich, że $||(A + B)x|| \leqslant C \cdot ||x||$, zatem

   $$||A + B|| \leqslant ||A|| + ||B||.$$

3. Było, że $A = 0 \Longrightarrow ||A|| = 0$. Pokażemy, że też: $||A = 0|| \Longrightarrow A = 0$.

   Dow. będzie niewprost: Pokażemy, że jeżeli $A \neq 0 \Longrightarrow ||A|| > 0$. Najsampierw zauważmy, że jeśli $A \neq 0$, to istnieje wektor x taki, że $Ax \neq 0$. Możemy założyć, że $||x|| = 1$. Niech $||Ax|| = k > 0$. Ponieważ $||Ax|| \leqslant ||A||$ dla każdego x takiego, że $||x|| = 1$, to znaczy, że $||A|| \geqslant k > 0$.

   CBDO w p. 3.

4. Mamy:

   $$(A \cdot B)x|| = ||A(Bx)|| \leqslant ||A|| \cdot ||Bx|| \leqslant ||A|| \cdot ||B|| \cdot ||x||;$$

   argumentując analogicznie jak pod koniec p. 2. mamy, że $||A \cdot B|| \leqslant ||A|| \cdot ||B||$.

Mamy następujący prosty fakt dotyczący funkcji rzeczywistych.

Najsampierw, przypomnijmy tw. Lagrange'a o wartości średniej:

Niech $f :]a,b[\to \mathbb{R}$ – różniczkowalna. Dla dowolnych $x,y \in]a,b[$ istnieje taki punkt $\xi \in]x,y[$, że

$$\frac{f(x)-f(y)}{x-y} = f'(\xi).$$

Teraz wzmiankowane

Stw. Niech pochodna $f'(x)$ będzie ograniczona na odcinku $]a,b[$: $|f'(x)| < C$ dla dowolnego $x \in]a,b[$. Wtedy dla dowolnych $x,y \in]a,b[$ mamy

$$|f(x)-f(y)| = \leqslant C|x-y|.$$

Dow. Bo: Dla dowolnych $x,y \in]a,b[$ zachodzi tw. Lagrange'a:

$$|f(x)-f(y)| = |f'(\xi)| \cdot |x-y|$$

gdzie ξ zależy od punktów x,y. Ale pochodna jest z założenia ograniczona przez C, więc powyższe oszacowanie zachodzi dla *dowolnych* punktów x,y:

$$|f(x)-f(y)| = |f'(\xi)| \cdot |x-y| \leqslant C|x-y|.$$

CBDO

Będziemy potrzebowali rozszerzenia tego faktu na odwzorowania. Okazuje się, że zachodzi

Stw. Niech $\mathcal{O} \subset \mathbb{R}^n$ – otwarty i wypukły. Niech $T : \mathcal{O} \to \mathcal{U} \subset \mathbb{R}^m$. Niech *norma* z pochodnej T będzie ograniczona, tzn. dla każdego $x \in \mathcal{O}$ niech $||T'(x)|| \leqslant C$. Wtedy

$$d(T(x),T(y)) \leqslant Cd(x,y)$$

dla dowolnych $x,y \in \mathcal{O}$.

Uzupełnienie Zbiór $X \subset \mathbb{R}^N$ nazywamy wypukłym, gdy dla dowolnych jego punktów x,y, także punkt $z = \alpha x + (1-\alpha)y$ należy do X dla dowolnego $\alpha \in [0,1]$. **RYS.**

. ? W tym miejscu dow. że kula jest wypukła?

Dow. Weźmy $h \in \mathbb{R}^n$ na tyle małe, aby $y = x + h$ należało do $\mathcal{O}$. (**RYS.**). Niech $\mathbb{R}^m \ni k = T(x+h) - T(x)$; naówczas $||k|| = d(T(x+h), T(x)$.

Zdefiniujmy funkcję zmiennej rzeczywistej λ, $\lambda \in [0,1]$:

$$f(\lambda) = \sum_{i=1}^{m} k^i (T^i(x+\lambda h) - T^i(x)).$$

Policzmy pochodną $f(\lambda)$:

$$\frac{\mathsf{d}f}{\mathsf{d}\lambda} = \sum_{i=1}^{m} k^i \sum_{j=1}^{n} \frac{\partial T^i}{\partial x^j}(x+\lambda h)h^j \tag{4}$$

Mamy: $f(0) = 0$ oraz $f(1) = ||k||^2$. Wobec tego, z tw. Lagrange'a o wartości średniej wnosimy, iż istnieje $\xi \in]0,1[$ t. że $f(1) - f(0) = f'(\xi)$. Pisząc jawnie wyrażenie (4) na $f'(\lambda)$ uzyskane wyżej, mamy

$$\left| \sum_{i=1}^{m} k^i \sum_{j=1}^{n} \frac{\partial T^i}{\partial x^j}(x+\xi h)h^j \right| = ||k||^2$$

Z własności 4. dla normy macierzy, mamy, iż lewa strona jest nie większa niż

$$||k|| \cdot ||T'(x+\xi h)h||$$

zatem

$$||k||^2 \leqslant ||k|| \cdot ||T'(x+\xi h)h|| \leqslant ||k|| \cdot ||T'(x+\xi h)|| \cdot ||h|| \leqslant ||k|| \cdot C \cdot ||h||,$$

zatem

$$||k|| \leqslant C \cdot ||h||,$$

co znaczy, że

$$d(T(x+h), T(x)) \leqslant C \cdot d(x+h, x)$$

czyli

$$d(T(x), T(y)) \leqslant C \cdot d(x, y)$$

CBDO

Za jakiś czas przyda nam się definicja:
Def. Mówimy, że odwzorowanie $T : X \to$ jest *zbliżające*, jeżeli

$$\forall x, y \in X : d(T(x), T(y)) < d(x, y).$$

Morał: Jeśli $C < 1$, to T jest zbliżające.
Teraz już jesteśmy gotowi, aby sformułować

1.3 Twierdzenie o lokalnej odwracalności

Twierdzenie o lokalnej odwracalności. **RYS.**
Niech $F : \mathcal{O} \to \mathbb{R}^n$, gdzie $\mathcal{O} \subset \mathbb{R}^n$. Zakładamy, że F jest klasy C^1. Niech $x_0 \in \mathcal{O}$ i niech $F'(x_0)$ będzie odwracalne, tzn. $\det F'(x_0) \neq 0$.

Wtedy istnieje otoczenie punktu x_0 (a więc i $K(x_0, r)$, $r > 0$) i istnieje otoczenie $\mathcal{U}$ punktu $F(x_0)$ takie, że odwzorowanie F obcięte do $K(x_0, r)$: $F|_{K(x_0,r)} : K(x_0, r) \to \mathcal{U}$ jest odwracalne. Odwzorowanie odwrotne do niego jest też klasy C^1.

Oznaczmy: $F|_{K(x_0,r)} = \tilde{F}$. Wtedy wyrażenie na pochodną $(\tilde{F}^{-1})'$ odwzorowania odwrotnego jest dane przez

$$(\tilde{F}^{-1})'(\tilde{F}(x)) = (\tilde{F}'(x))^{-1}. \tag{5}$$

Dow. Jak pamiętamy, odwzorowanie odwrotne jest zdefiniowane przez: $\tilde{F}^{-1} \circ \tilde{F} = \mathrm{Id}$ lub, jawnie wypisując argument(-y),

$$(\tilde{F}^{-1})(\tilde{F}(x)) = x \tag{6}$$

Jeśli $\tilde{F}$ oraz $\tilde{F}^{-1}$ są klasy C^1, to obliczając pochodne obu stron wyrażenia (6) i korzystając z wzoru na pochodną odwzorowania złożonego mamy

$$(\tilde{F}^{-1})'(\tilde{F}(x)) \cdot (\tilde{F}'(x)) = I,$$

(I – macierz jednostkowa) czyli

$$(\tilde{F}^{-1})'(\tilde{F}(x)) = (\tilde{F}'(x))^{-1}.$$

czyli mamy wzór (5). Pozostaje wykazać całą resztę tezy.

Oznaczmy: $F'(x_0) = A$ i wybierzmy $\lambda \in \mathbb{R}_+$ tak, aby $4\lambda||A^{-1}|| = 1$. Ponieważ F jest klasy C^1, tzn. jego pochodne cząstkowe są ciągłe, to istnieje kula otwarta U o środku w punkcie x_0 taka, że

$$||F'(x) - A|| < 2\lambda \quad \text{dla wszystkich} \quad x \in U. \tag{7}$$

Załóżmy, że $x \in U$, $x + h \in U$ i zdefiniujmy $\Phi : [0, 1] \to \mathbb{R}^n$:

$$\Phi(t) = F(x + th) - F(x) - tAh \qquad 0 \leqslant t \leqslant 1.$$

Ponieważ kula U jest zbiorem wypukłym[2], to $x + th \in U$ dla $0 \leqslant t \leqslant 1$ i z (7) wynika, że

$$||\Phi'(t)|| = ||F'(x+th)h-Ah|| = ||(F'(x+th)-A)h|| \leqslant ||F'(x+th)-A||\cdot||h|| \leqslant 2\lambda||h|| \leqslant \frac{1}{2}||Ah|| \tag{8}$$

(w przedostatniej nierówności skorzystaliśmy z (7)). Ostatnia nierówność wynika z następującej argumentacji:

$$2\lambda||h|| = 2\lambda||A^{-1}Ah|| \leqslant 2\lambda||A^{-1}|| \cdot ||Ah|| = \frac{1}{2}||Ah||. \tag{9}$$

Przypomnijmy sobie teraz Stw. ze str. 6 mówiące, że jeżeli odwzorowanie $T : \mathbb{R}^m \to \mathbb{R}^n$ ma pochodną ograniczoną przez C, to dla dowolnych $x, y \in \mathbb{R}^m$ zachodzi nierówność $d(T(x), T(y)) \leqslant Cd(x, y)$. Zastosujmy go do funkcji $\Phi : \mathbb{R} \to \mathbb{R}^n$: Jeżeli jest spełnione oszacowanie (8) na normę Φ, to biorąc $x = 1, y = 0$ dostajemy

$$||\Phi(1) - \Phi(0)|| \leqslant \frac{1}{2}||Ah||,$$

co można przepisać jako

$$||F(x + h) - F(x) - Ah|| \leqslant \frac{1}{2}||Ah||. \tag{10}$$

Oznaczmy na chwilę $F(x + h) - F(x) = \Delta$. Mamy więc: $||\Delta - Ah|| \leqslant \frac{1}{2}||Ah||$ lub

$$-||\Delta - Ah|| \geqslant -\frac{1}{2}||Ah||; \tag{11}$$

Ponadto:

$$||Ah|| = ||\Delta - Ah - \Delta|| \leqslant ||\Delta - Ah|| + ||\Delta||,$$

czyli

$$||\Delta|| \geqslant ||Ah|| - ||\Delta - Ah||$$

zaś uwzględniając (11) mamy:

$$||\Delta|| \geqslant ||Ah|| - \frac{1}{2}||Ah|| = \frac{1}{2}||Ah||,$$

czyli, uwzględniając jeszcze (9), dostajemy

$$||F(x + h) - F(x)|| \geqslant \frac{1}{2}||Ah|| \geqslant 2\lambda||h||. \tag{12}$$

[2] To chyba nie było dowodzone; wydaje się, że warto

Nierówności (10) i (12) zachodzą dla *dowolnych* x i h takich, że $x \in U$ i $x + h \in U$. Tak więc nierówność (12) mówi, że F jest *wzajemnie jednoznaczna* na U (bowiem nie ma takich punktów $x, x+h$ aby zachodziło $F(x) = F(x+h)$).

Pozostaje pokazać *ciągłość* i *różniczkowalność* odwzorowania odwrotnego. Oznaczmy odwzorowanie odwrotne do F przez G.

Niech $\mathcal{U} = F(U)$, niech $y \in \mathcal{U}$, $y + k \in V$ i niech $x = G(y)$. Niech

$$h = G(y+k) - G(y).$$

Pamiętamy, że na U pochodna $F'(x)$ ma operator odwrotny, który oznaczymy przez B.

Odwzorowanie F jest różniczkowalne, więc możemy zapisać:

$$k = F(x+h) - F(x) = F'(x)h + r(h),$$

gdzie $r(h)$ jest *resztą*, tzn. zachodzi: $\frac{||r(h)||}{||h||} \to 0$ dla $h \to 0$. Na obie strony powyższej równości zadziałajmy operatorem B. Otrzymamy: $Bk = h + Br(h)$ lub

$$G(y+k) - G(y) = Bk - B(r(h)). \tag{13}$$

Na mocy (12), $2\lambda||h|| \leqslant ||k||$. Zatem $h \to 0$, jeśli $k \to 0$ (co dowodzi ciągłości G w punkcie y) oraz

$$\frac{||B(r(h))||}{||k||} \leqslant \frac{||B||}{2\lambda} \cdot \frac{||r(h)||}{||h||} \to 0, \quad \text{gdy} \quad k \to 0. \tag{14}$$

Z porównania (13) i (14) wynika, że G jest różniczkowalna w punkcie y oraz że $G'(y) = B$. Można to przeformułować mówiąc, że dla $y \in \mathcal{U}$ zachodzi

$$G'(y) = [F'(G(y))]^{-1}. \tag{15}$$

o czym już wiedzieliśmy z formalnego rachunku tuż przed dowodem (ale dopiero teraz uzasadniliśmy poprawność tego rachunku).

CBDO

Przykł: Dla $F : \mathbb{R}^2 \ni (r, \phi) \to (x, y) \in \mathbb{R}^2$ określonego jako: $x(r, \phi) = r\cos\phi$, $y(r, \phi) = r\sin\phi$ pokazujemy bezpośrednim rachunkiem, że $(F^{-1})' = (F')^{-1}$.

Konkretnie: Liczymy LHS, RHS i porównujemy (kiedyś część już liczyliśmy, ale niech teraz rachunki będą w jednym miejscu).

RHS

$$F' \equiv DF = \begin{bmatrix} x_r & x_\phi \\ y_r & y_\phi \end{bmatrix} = \begin{bmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{bmatrix}$$

Pamiętamy (? – jełsi ktoś nie pamięta to trudno, niech wyprowadzi albo chociaż sprawdzi) że macierzą odwrotną do dowolnej macierzy dwuwymiarowej $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ jest

$$A^{-1} \equiv \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Tu mamy:

$$\det(DF) = r,$$

i na całą macierz odwrotną

$$(DF)^{-1} = \frac{1}{r}\left[\begin{array}{cc} r\cos\phi & r\sin\phi \\ -\sin\phi & \cos\phi \end{array}\right]. \tag{16}$$

A teraz!!

RHS:

A teraz znajdziemy odwzorowanie *odwrotne* do F i obliczymy jego pochodną. Aby znaleźć F^{-1}, musimy wyrazić r, ϕ przez x, y. Robiliśmy to już kiedyś, jest to proste, więc tylko przypomnimy:

$$r = \sqrt{x^2+y^2}, \qquad \phi = \operatorname{arctg}\frac{y}{x}.$$

Policzmy niezbędne pochodne, wyrażając wynik w zmiennych (r, ϕ):

$$r_x = \frac{2x}{2\sqrt{x^2+y^2}} = \frac{r\cos\phi}{r} = \cos\phi; \qquad r_y = \frac{2y}{2\sqrt{x^2+y^2}} = \frac{r\cos\phi}{r} = \sin\phi.$$

$$\phi_x = \frac{-\frac{y}{x^2}}{1+\frac{y^2}{x^2}} = -\frac{y}{x^2+y^2} = -\frac{r\sin\phi}{r^2} = -\frac{1}{r}\sin\phi;$$

$$\phi_y = \frac{\frac{1}{x}}{1+\frac{y^2}{x^2}} = \frac{x}{x^2+y^2} = \frac{r\cos\phi}{r^2} = \frac{1}{r}\cos\phi.$$

Układając je w macierz, otrzymamy:

$$D(F^{-1}) = \left[\begin{array}{cc} r_x & r_y \\ \phi_x & \phi_y \end{array}\right] = \left[\begin{array}{cc} \cos\phi & \sin\phi \\ -\frac{1}{r}\sin\phi & \frac{1}{r}\cos\phi \end{array}\right] \tag{17}$$

Porównując (16) i (17) widać, że otrzymaliśmy to samo – jak i powinno być.

2 Tw. o funkcji uwikłanej

Aby wyrobić intuicję, rozpatrzmy najsampierw układy równań *liniowych*. Taki układ to m równań na N zmiennych, gdzie założymy, że $N > m$. [3] W takiej sytuacji, jeśli jest spełniony określony warunek, który zaraz wypiszemy, możemy wyrazić m zmiennych jako funkcję pozostałych $N - m$.

Przykłady.

1. $N = 3$, $m = 1$. Weźmy równanie:

   $$3x + 2y + z = 1$$

   (Geometrycznie, powyższe równanie opisuje *płaszczyznę* w $\mathbb{R}^3$ – więc obiekt dwuwymiarowy.) Jedną ze zmiennych (np. z) można wyrazić jako funkcję od pozostałych dwóch x, y:

   $$z = 1 - 3x - 2y$$

[3] Dlaczego zakładamy, że ilość równań jest mniejsza od ilości niewiadomych? Bo gdy jest większa, tzn. $N < m$, to – jeśli równania są liniowo niezależne – to układ nie ma rozwiązań, a gdy $N = m$, to mamy sytuację z tw. o lokalnej odwracalności.

2. $N = 3$, $m = 2$. Weźmy układ 2 równań na 3 niewiadome:

$$\begin{cases} x + y + z &= 1 \\ x + 2y - z &= 1 \end{cases}$$

(Geometrycznie, powyższy układ dwóch równań opisuje przecięcie dwóch płaszczyzn, a więc *prostą*). Wybierzmy dwie zmienne, np. x, y i wyraźmy je jako funkcje pozostałej zmiennej z. Mamy:

$$W = \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = 1, \quad W_x = \begin{vmatrix} 1-z & 1 \\ 1+z & 2 \end{vmatrix} = 1 - 3z, \quad W_y = \begin{vmatrix} 1 & 1-z \\ 1 & 1+z \end{vmatrix} = 2z$$

czyli rozwiązaniem jest: $\begin{cases} x &= 1 - 3z \\ y &= 2z \end{cases}$

3. W ogólnym przypadku m równań na N zmiennych, wybieramy m zmiennych które chcemy wyrazić jako funkcje $N - m$ pozostałych. Zmienne zalezne przenosimy na *lewą* stronę układu, a zmienne niezależne – na prawą, traktując je jako parametry. Układ da się rozwiązać, jeśli główny wyznacznik jest różny od zera.

Wróćmy teraz do sytuacji, którą będziemy chcieli analizować: Będzie to układ m równań na N zmiennych, ale równań na ogół *nieliniowych*.

W ogólnym przypadku rozwiązywanie takich układów jest bardzo trudne (podobnie jak przy konstrukcji odwzorowania odwrotnego). Jeżeli jednak ograniczymy się do sytuacji *lokalnych*, tzn. małego otoczenia jakiegoś punktu z $\mathbb{R}^N$, to sytuacja pod wieloma względami przypomina to, z czym mamy do czynienia w przypadku układów równań liniowych. Zanim sformułujemy odpowiednie twierdzenie, podeprzemy się znów dwoma przykładami.

1. $N = 2$, $m = 1$.

$$x^2 + y^2 = 2$$

co możemy zapisać jako: $H(x, y) = 0$, gdzie $H(x, y) = x^2 + y^2 - 2$. Weźmy punkt $p = (1, 1)$ na płaszczyźnie; widać, że w otoczeniu tego punktu można wyrazić jedną ze zmiennych, np. y jako funkcję pozostałej (tu x). **RYS.** (Rozwiązanie można tu napisać jawnie, tzn. $y = +\sqrt{2 - x^2}$). Inna jest sytuacja w otoczeniu punktu $p^* = (2, 0)$. W żadnym otoczeniu tego punktu nie można *jednoznacznie* wyrazić y jako funkcji x.

2. $N = 3$, $m = 2$.

$$W = \begin{cases} x^2 + y^2 + z^2 - 1 &= 0 \\ x + 2y + 3z &= 0 \end{cases}$$

Co opisuje ten układ równań? Pierwsze równanie to równanie *sfery*, a drugie – *płaszczyzny*, zatem powyższy układ to *przecięcie* sfery z płaszczyzną, czyli *okrąg* (łatwo się przekonać, że nie jest to zbiór pusty ani punkt). Rozwiązaniem powyższego układu byłaby para zmiennych, np. x i y jako funkcja pozostałej trzeciej: $\begin{cases} x &= x(z), \\ y &= y(z) \end{cases}$, czyli opis parametryczny okręgu. Jest to możliwe dla prawie wszystkich punktów okręgu, z wyjątkiem jednak niektórych z nich. Poniżej zobaczymy, jak rozpoznać, kiedy w otoczeniu danego punktu jest możliwy taki jednoznaczny opis parametryczny, a kiedy nie jest możliwy.

3. **Przykład z innej beczki – termodynamika**. Równanie stanu, np. $F(p, V, T) = 0$ i konieczność policzenia stąd np. $p(V, T)$

4. **Przykład z jeszcze innej beczki – mechanika**. Układy z więzami (np. punkt uwięziony na powierzchni i ślizgający się tylko po niej).

Rozpatrzmy teraz przypadek ogólny. Zmienimy najpierw trochę oznaczenia: Ponieważ $N > m$, będziemy pisać: $N = n+m$ (gdzie $n > 0$). Mamy zatem układ m równań na $n+m$ zmiennych. Będziemy ten układ (lokalnie) rozwiązywać, tzn. wyznaczać m zmiennych jako funkcje n pozostałych. Zmienne *niezależne* oznaczać będziemy jako $x = (x_1, \ldots, x_n)$, zaś zmienne *zależne* jako $y = (y_1, \ldots, y_m)$.

Załóżmy więc, że mamy odwzorowanie $H : \mathbb{R}^n \times \mathbb{R}^m \supset \mathcal{O} \to \mathbb{R}^m$ ($\mathcal{O}$ jest zb. otwartym) klasy C^1. $H(x,y)$ jest więc wektorem o m składowych:

$$H(x,y) = \begin{pmatrix} H^1(x,y) \\ H^2(x,y) \\ \vdots \\ H^m(x,y) \end{pmatrix}$$

zaś równość: $H(x,y) = 0$ możemy przepisać jako m równań:

$$H(x,y) = \begin{cases} H^1(x,y) &= 0 \\ H^2(x,y) &= 0 \\ & \vdots \\ H^m(x,y) &= 0 \end{cases}$$

Popatrzmy jeszcze na macierz pochodnej H'. Jest to macierz rozmiaru $m \times (n+m)$:

$$H' = \begin{bmatrix} \frac{\partial H^1}{\partial x^1} & \cdots & \frac{\partial H^1}{\partial x^n} & \frac{\partial H^1}{\partial y^1} & \cdots & \frac{\partial H^1}{\partial y^m} \\ \frac{\partial H^2}{\partial x^1} & \cdots & \frac{\partial H^2}{\partial x^n} & \frac{\partial H^2}{\partial y^1} & \cdots & \frac{\partial H^2}{\partial y^m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial H^m}{\partial x^1} & \cdots & \frac{\partial H^m}{\partial x^n} & \frac{\partial H^m}{\partial y^1} & \cdots & \frac{\partial H^m}{\partial y^m} \end{bmatrix} \tag{18}$$

Są tam pochodne po zmiennych x oraz y. Macierz pochodnych H po zmiennych x oznaczymy jako H'_x (jest to macierz $m \times n$), zaś po zmiennych y jako H'_y (jest to macierz $m \times m$). Możemy więc napisać

$$H' = (H'_x, H'_y)$$

Tw. (**o funkcji uwikłanej**). Niech $H : \mathbb{R}^n \times \mathbb{R}^m \supset \mathcal{O} \to \mathbb{R}^m$ ($\mathcal{O}$ jest zb. otwartym) będzie odwzorowaniem klasy C^1. Niech $H(x_0, y_0) = 0$. Niech $H'_y(x_0, y_0)$ będzie odwracalna.

Wtedy istnieje otoczenie $\mathcal{U}$ punktu x_0: $\mathcal{U} \subset \mathbb{R}^n$ oraz odwzorowanie ϕ klasy C^1: $\phi : \mathcal{U} \to \mathbb{R}^m$ takie, że

$$H(x, \phi(x)) \equiv 0 \quad \text{dla} \quad x \in \mathcal{U} \tag{19}$$

oraz pochodna $\phi'(x)$ jest równa

$$\phi'(x) = -(H'_y(x, \phi(x)))^{-1} \cdot (H'_x(x, \phi(x))). \tag{20}$$

Dow. Zdefiniujmy odwzorowanie Ψ następująco:

$$\Psi : \mathcal{O} \ni (x,y) \to (x, H(x,y)) \in \mathbb{R}^n \times \mathbb{R}^m,$$

czyli jawnie, w składowych:

$$\begin{pmatrix} x^1 \\ \vdots \\ x^n \\ y^1 \\ \vdots \\ y^m \end{pmatrix} \xrightarrow{\Psi} \begin{pmatrix} x^1 \\ \vdots \\ x^n \\ H^1(x,y) \\ \vdots \\ H^m(x,y) \end{pmatrix} \quad \text{co daje} \quad \Psi'(x_0, y_0) = \begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ H'_x(x_0, y_0) & H'_y(x_0, y_0) \end{pmatrix}$$

gdzie $\mathbf{I}_n$ jest macierzą jednostkową rozmiaru $n \times n$, $\mathbf{0}$ jest macierzą rozmiaru $n \times m$ złożoną z samych zer.

Mamy:

$$\det(\Psi'(x_0, y_0)) = \det(H'_y(x_0, y_0) \neq 0 \quad \text{z założenia.}$$

zatem – z twierdzenia o lokalnej odwracalności – istnieje otoczenie $\mathcal{V}$ punktu $(x_0, 0)$ oraz otoczenie $\mathcal{W}$ punktu (x_0, y_0) oraz istnieje odwzorowanie Ψ^{-1} określone na $\mathcal{V}$: $\Psi^{-1} : \mathcal{V} \to \mathcal{W}$ takie, że

$$\Psi^{-1}(x, z) = (x, \tilde{\phi}(x, z)) \in \mathcal{W}$$

Odwzorowanie $\tilde{\phi}$ jest klasy C^1. Oznaczmy teraz:

$$\phi(x) = \tilde{\phi}(x, 0);$$

mamy:

$$\Psi(x, \tilde{\phi}(x, z)) = (x, z)$$

i z definicji odwzorowania Ψ

$$\Psi(x, \tilde{\phi}(x, z)) = (x, H(x, \tilde{\phi}(x, z))) = (x, z)$$

i patrząc na drugie składowe powyższej równości dla $z = 0$ mamy

$$H(x, \phi(x)) = H(x, \tilde{\phi}(x, 0)) = 0$$

Znaleźliśmy więc odwzorowanie ϕ o własnościach danych przez (19).

Co do wzoru (20) na pochodną, to rozważmy następujące odwzorowanie $F : \mathbb{R}^n \to \mathbb{R}^m$:

$$F(x) = H(x, \phi(x)).$$

F jest odwzorowaniem tożsamościowo równym zeru, więc jego pochodna też jest tożsamościowo równa zeru (i wyższe pochodne też). Policzmy pochodną F':

?? Bardziej szczegółowa kalkulacja??

$$F'(x) = H'_x(x, \phi(x) + H'_y(x, \phi(x)) \cdot \phi'(x) \equiv 0,$$

co daje

$$-H'_x(x, \phi(x) = H'_y(x, \phi(x)) \cdot \phi'(x)$$

i po pomnożeniu (lewostronnym) przez macierz $(H'_y(x, \phi(x)))^{-1}$ (a pomnożyć można, bo w dostatecznie małym otoczeniu x_0 macierz $H'_y(x, \phi(x))$ jest odwracalna) dostajemy wzór (20).

CBDO

Przykł. $H : \mathbb{R}^3 \to \mathbb{R}$; czyli $m = 1, n = 2$; czyli mamy tu jedno równanie na 3 zmienne: $H(x, y, z) = 0$ i chcemy stąd wyrazić z jako funkcję od pozostałych zmiennych x, y: $z = z(x, y)$ w otoczeniu jakiegoś danego punktu (x_0, y_0, z_0). Udowodnione dopiero co twierdzenie o funkcji uwikłanej mówi, że jest to możliwe, gdy pochodna $\frac{\partial H}{\partial z}(x_0, y_0, z_0) \neq 0$. Gdybyśmy jeszcze chcieli policzyć pochodne z po swoich argumentach, to są one następujące:

$$\frac{\partial z}{\partial x} = -\frac{\frac{\partial H}{\partial x}}{\frac{\partial H}{\partial z}}, \quad \frac{\partial z}{\partial y} = -\frac{\frac{\partial H}{\partial y}}{\frac{\partial H}{\partial z}}.$$