

Tavelpresentation Flervariabelanalys

Leo Andersson, Liam Antonsson, Gustaf Fransén Björn,
Vincent Harbander, Samuel Kronquist, Mathias Ohlsson

1 februari 2022

1 Sammanfattning läsvecka 2

- Taylors formel - Vincent
- Undersökning av lokala extrema - Gustaf
- Vektorvärda funktioner - Samuel
- Funktionalmatris och funktionaldeterminant - Liam
- Linjärisering och kedjeregeln i allmänhet - Mathias
- Implicita funktionssatsen - Leo

2 Taylors formel

Taylors formel i flera variabler är en utvidgning av Taylors formel i en variabel. Funktionen skrivs som ett polynom och en restterm.

$$f(a+h) = P_k(a, h) + Q_{k+1}(a, h)$$

Låt D vara omgivningen kring $x=a$.

$$f \in C^{k+1}(D)$$

In en variabel får man formeln

$$P_k(a, h) = \sum_{j=0}^k \frac{f^{(j)}(a)}{(j)!} h^j,$$

$$Q_{k+1}(a, h) = \frac{f^{(k+1)}(\xi_h)}{(k+1)!} h^{k+1},$$

där $\xi_h \in [a - |h|, a + |h|]$ Resttermens exakta form är vanligtvis inte intressant, utan det intressanta är att den har egenskapen

$$Q_{k+1} = \mathcal{O}(h^{k+1}).$$

Idén med utvidgningen till fler variabler är att göra en endimensionell omskrivning av funktionen f . Låt $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ där vi vill utveckla kring $(a, b) \subseteq D \subseteq \mathbb{R}^2$.

$$f \in C^{k+1}(D)$$

Fixera två tal h_1 och h_2 .

$$F(t) = f(a + th_1, b + th_2)$$

Vi vill Taylorutveckla F kring $t = 0$. Antag $t > 0$.

$$F(t) = \sum_{j=0}^k \frac{F^{(j)}(0)}{(j)!} t^j + \frac{F^{(k+1)}(\xi_t)}{(k+1)!} t^{k+1}$$

för något $\xi_t \in (0, t)$ Vi vill uttrycka derivatorna av F uttryckt med partiella derivator av f . Vi använder kedjeregeln.

$$F'(t) = f'_x(a+th_1, b+th_2) \frac{d}{dt}(a+th_1) + f'_y(a+th_1, b+th_2) \frac{d}{dt}(b+th_2) = (h_1 \frac{\partial}{\partial x} + h_2 \frac{\partial}{\partial y}) f(a+th_1, b+th_2)$$

Observera att relationen mellan F' och f är samma som mellan F och f . Denna relation kan användas iterativt för att få högre derivator av F .

$$F^{(j)}(t) = (h_1 \frac{\partial}{\partial x} + h_2 \frac{\partial}{\partial y})^j f(a + th_1, b + th_2)$$

Minns att h_1 och h_2 är fixerade. Det innebär att de är konstanter. Att $f \in C^{k+1}(D)$ innebär att $\frac{\partial}{\partial x}$ och $\frac{\partial}{\partial y}$ kommuterar enligt Clairouts sats.

Nu kan binomialsatsen användas för att utveckla parentesen generellt och skriva det som en summa.

$$F^{(j)}(t) = \sum_{i=0}^j \binom{j}{i} h_1^i h_2^{j-i} \frac{\partial^j f(a + th_1, b + th_2)}{\partial x^i \partial y^{j-i}}$$

Med vektornotation kan man skriva detta mer kortfattat. Låt $\mathbf{a} = (a, b)$, $\mathbf{h} = (h_1, h_2)$ och $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$.

$$h_1 \frac{\partial}{\partial x} + h_2 \frac{\partial}{\partial y} = \mathbf{h} \cdot \nabla$$

Detta innebär att

$$F^{(j)}(t) = (\mathbf{h} \cdot \nabla)^{(j)} f(\mathbf{a} + t\mathbf{h}).$$

När man använder Taylors formel får man till slut

$$f(\mathbf{a} + \mathbf{h}) = F(1) = \sum_{j=0}^k \frac{(\mathbf{h} \cdot \nabla)^j}{j!} f(\mathbf{a}) + \frac{(\mathbf{h} \cdot \nabla)^{k+1}}{(k+1)!} f(\xi_t)$$

där ξ_t är en vektor på sträckan mellan $\mathbf{a}$ och $\mathbf{a} + \mathbf{h}$.

Den grafiska tolkningen för utvidgningmetoden är att vi Taylorutvecklar längs en linje som specificeras av h_1 och h_2 som passerar genom $\mathbf{a}$. h_1 och h_2 kan väljas för att skapa varje linje i planet. Parametriseringen för linjen är

$$x = a + th_1 \quad y = b + th_2.$$

3 Undersökning av lokala extrema

En lokal extrempunkt för en funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ är antingen ett lokalt maximum eller minimum. För ett maximum så gäller det att f har ett lokalt maximum i $\vec{a} \in \mathbb{R}^n$ om det finns en omgivning Ω av $\vec{a}$ så att $f(\vec{x}) \leq f(\vec{a}), \forall \vec{x} \in \Omega$. Om $\vec{a}$ är ett minimum gäller istället $f(\vec{x}) \geq f(\vec{a}), \forall \vec{x} \in \Omega$. För stränga lokala maximum och minimum ersätts $\leq$ och $\geq$ med $<$ och $>$, och mängden ändras till $\forall \vec{x} \in \Omega \setminus \{\vec{a}\}$, annars definieras de på samma sätt. En lokal extrempunkt är alltså en lokal max- eller minpunkt.

Om $f(\vec{x})$ har en lokal extrempunkt i en inre punkt $\vec{a}$ i definitionsmängden och f är partiellt deriverbar i $\vec{a}$ så är $\nabla f(\vec{a}) = \vec{0}$. Det vill säga att de partiella derivatorna av grad ett i $\vec{a}$ är 0. Viktigt att notera är att omvändningen **inte** gäller. Är $\nabla f(\vec{a}) = \vec{0}$, så kallas $\vec{a}$ en stationär punkt, vilket inte nödvändigtvis är en extrempunkt.

För att studera stationära punkter och ta reda på deras klassificering, är Taylorutvecklingen till god hjälp. Antag att vi har en funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ med en stationär punkt i $\vec{a} = \begin{bmatrix} a \\ b \end{bmatrix}$ (det fungerar också i fler dimensioner, men då ser Taylorpolynomet annorlunda ut). Taylorutveckling till grad två leder till

$$f(a+h, b+k) - f(a+b) = \frac{1}{2}(Ah^2 + 2Bhk + Ck^2) + \mathcal{O}\left(\left\|\begin{bmatrix} h \\ k \end{bmatrix}\right\|^3\right)$$

där $A = f''_{xx}(\vec{a})$, $B = f''_{xy}(\vec{a})$ och $C = f''_{yy}(\vec{a})$, då $\nabla f(\vec{a}) = \vec{0}$ i stationära punkter. Tecknet på $f(a+h, b+k) - f(a+b)$ för små $\begin{bmatrix} h \\ k \end{bmatrix}$ avgör den stationära punktens klassificering. Låt $Q(h, k) = Ah^2 + 2Bhk + Ck^2$. $Q(h, k)$ är då $\mathcal{O}\left(\left\|\begin{bmatrix} h \\ k \end{bmatrix}\right\|^2\right)$ och dominerar **oftast** $\mathcal{O}\left(\left\|\begin{bmatrix} h \\ k \end{bmatrix}\right\|^3\right)$ -termen.

Tecknet på $Q(h, k)$ kommer alltså avgöra klassificeringen av den stationära punkten, och $Q(\vec{x})$ kallas **kvadratisk form**. En kvadratisk form $Q(\vec{x})$ är:

- Positivt definit om $Q(\vec{x}) > 0, \forall \vec{x} \neq \vec{0}$ (positivt semidefinit om $Q(\vec{x}) \geq 0$).
Är Q positivt definit är $\vec{a}$ strängt lokalt minimum.
- Negativt definit om $Q(\vec{x}) < 0, \forall \vec{x} \neq \vec{0}$ (negativt semidefinit om $Q(\vec{x}) \leq 0$).
Är Q negativt definit är $\vec{a}$ strängt lokalt maximum.
- Indefinit om $Q(\vec{x})$ antar båda positiva och negativa värden. Är Q indefinit är $\vec{a}$ en sadelpunkt.

Är Q semidefinit så måste Taylorutvecklingen förlängas, det säger inget om den stationära punktens klassificering.

Klassificeringen för en stationär punkt kan också bestämmas med hjälp av matriser. Låt $M = \begin{pmatrix} f''_{xx} (= A) & f''_{xy} (= B) \\ f''_{xy} (= B) & f''_{yy} (= C) \end{pmatrix}$ i punkten $\vec{a}$, och $\vec{h} = \begin{bmatrix} h \\ k \end{bmatrix}$. Det kan utökas och gäller i fler dimensioner också. M är då en symmetrisk $n \times n$ -matris

med de partiella derivatorna i $\vec{a}$ av grad två enligt mönstret för två dimensioner, och $\vec{h} \in \mathbb{R}^n$, och $Q(\vec{h}) = \vec{h}^T M \vec{h}$. Egenvärdena är rötterna till den karakteristiska ekvationen $\det(M - \lambda I) = 0$. Då är $Q(\vec{h}) = \vec{h}^T M \vec{h}$

- Positivt definit om alla egenvärden är positiva.
- Negativt definit om alla egenvärden är negativa.
- Indefinit om M har både positiva och negativa egenvärden.
- Semidefinit om något egenvärde är 0.

I specialfallet två dimensioner kan matrisformuleringen förenklas till att om $\det(M) = AC - B^2$

- > 0 och $A > 0$ så är $Q(\vec{x})$ positivt definit.
- > 0 och $A < 0$ så är $Q(\vec{x})$ negativt definit.
- < 0 så är $Q(\vec{x})$ indefinit.
- $= 0$ så är $Q(\vec{x})$ semidefinit.

Klassificeringen av den stationära punkten $\vec{a}$ avgörs precis på samma sätt som tidigare av sorten på Q även med matrismetoderna.

4 Vektorvärda funktioner

Enklare formen av en vektorvärd funktion:

$$\mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}^p$$

Som skrivs:

$$\mathbf{f}(x) = (f_1(x), \dots, f_p(x)),$$

där varje f_i är en skalärvärd funktion av 1 variabel.

Om vi istället kollar på en vektorvärd funktion av följande form:

$$\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^p$$

Som skrivs:

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_p(\mathbf{x})),$$

där $\mathbf{x} = [x_1, \dots, x_n]$ och f_i är en skalärvärd funktion av n variabler. Denna typen av vektorvärda funktioner är väldigt användbara och speciellt användbart är specialfallet då $p = n$ som används bl.a vid koordinatbyten.

Differentierbarhet för en vektorvärd funktion i $\mathbf{a} := [a_1, \dots, a_n]^T$ gäller om och endast om $f_i(\mathbf{x})$ är differentierbar i punkten $\mathbf{a}$, $\forall i = 1, \dots, p$.

Om detta gäller går det att linjärisera funktionen på följande form:

$\exists \rho_i : \mathbb{R}^n \rightarrow \mathbb{R}$, så att

$$f_i(\mathbf{a} + \mathbf{h}) = f_i(\mathbf{a}) + \nabla f'_i(\mathbf{a})\mathbf{h} + \|\mathbf{h}\|\rho_i(\mathbf{h}),$$

där $\rho_i(\mathbf{h}) \rightarrow 0$ då $\mathbf{h} \rightarrow \mathbf{0}$

5 Funktionalmatris och funktionaldeterminant

Funktionalmatrisen definieras som

$$\frac{\partial(\mathbf{f})}{\partial(\mathbf{x})} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix},$$

där $\mathbf{f}$ är en vektorvärd funktion från $\mathbb{R}^n$ till $\mathbb{R}^m$ och $\mathbf{x} \in \mathbb{R}^n$. Med funktionalmatrisen får man en komplett bild över $\mathbf{f}$:s beroende med avseende på alla variabler, varför den ibland kallas derivatan av $\mathbf{f}$. En annan mer kompakt form av funktionalmatrisen är

$$\frac{\partial(\mathbf{f})}{\partial(\mathbf{x})} = \begin{bmatrix} \nabla f_1^T \\ \nabla f_2^T \\ \vdots \\ \nabla f_m^T \end{bmatrix}.$$

På denna form åskådliggörs det att radvektorerna i funktionalmatrisen är ∇f_i^T . Gradienten måste transponeras så att den blir en radvektor istället för en kolonnvektor. På liknande sätt utgör kolonnvektorerna den partiella derivatan av $\mathbf{f}$ med avseende på något x_i . En fördel med ovanstående skrivsätt är att det går smidigare att skriva kedjeregeln av $\mathbf{f}$, från $\mathbb{R}^n$ till $\mathbb{R}^m$, på $\mathbf{g}$, från $\mathbb{R}^p$ till $\mathbb{R}^n$, om $\mathbf{x} = \mathbf{g}(\mathbf{t})$ för $\mathbf{t} \in \mathbb{R}^p$

$$\frac{\partial(\mathbf{f} \circ \mathbf{g})}{\partial(\mathbf{t})} = \begin{bmatrix} \nabla f_1^T \\ \nabla f_2^T \\ \vdots \\ \nabla f_m^T \end{bmatrix} \begin{bmatrix} \nabla g_1^T \\ \nabla g_2^T \\ \vdots \\ \nabla g_n^T \end{bmatrix}. \quad (1)$$

Riktigheten i uttrycket kan lätt kontrolleras genom att jämföra exempelvis första elementet i V.L med den "vanliga" kedjeregeln i flervariabeln när $\mathbf{f}$ inte är vektorvärd.

Funktionaldeterminanten är helt enkelt determinanten av funktionalmatrisen och skrivs på liknande sätt som för funktionalen fast med d istället för ∂

$$\det\left(\frac{\partial(\mathbf{f})}{\partial(\mathbf{x})}\right) = \frac{d(\mathbf{f})}{d(\mathbf{x})} = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{vmatrix}.$$

Eftersom determinanten är distributiv gäller det från ekvation (5) att

$$\frac{d(\mathbf{f} \circ \mathbf{g})}{d(\mathbf{t})} = \frac{d(\mathbf{f})}{d(\mathbf{x})} \cdot \frac{d(\mathbf{g})}{d(\mathbf{t})}.$$

6 Linjärisering och kedjeregeln i allmänhet

Linjärisering för en funktion från $\mathbb{R}^n$ till $\mathbb{R}^p$. Linjärisering kan ses som en Taylorutveckling utan resttermen.

Formeln för linjärisering är:

$$\mathbf{f}(\mathbf{a} + \mathbf{h}) = \mathbf{f}(\mathbf{a}) + \mathbf{f}'(\mathbf{a})\mathbf{h} + \|\mathbf{h}\|\boldsymbol{\rho}(\mathbf{h}) \quad (2)$$

Kravet på $\boldsymbol{\rho}$ är att när $\mathbf{h}$ går mot $\mathbf{0}$ så kommer $\boldsymbol{\rho}$ gå mot $\mathbf{0}$. För själva linjärisering är hela formeln förutom den sista termen.

Relevant notation:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix} \quad \mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \dots \\ h_n \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \quad \mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \dots \\ f_p(\mathbf{x}) \end{bmatrix} \quad \boldsymbol{\rho}(\mathbf{h}) = \begin{bmatrix} \rho_1(\mathbf{h}) \\ \rho_2(\mathbf{h}) \\ \dots \\ \rho_p(\mathbf{h}) \end{bmatrix}$$

Vi har redan sett kedjeregeln är flera variabler men med olika begränsningar. Den allmänna kedjeregeln för flervariabler.

Kedjeregeln i allmänhet mellan två funktioner $\mathbf{g}: \mathbb{R}^n$ till $\mathbb{R}^p$ och $\mathbf{f}: \mathbb{R}^p$ till $\mathbb{R}^m$. Det antas att båda funktionerna är deriverbara en gång. Vi antar också att det finns en baspunkt $\mathbf{t}$ som tillhör $\mathbb{R}^n$ vilket är en inre punkt till definitionsmängden till $\mathbf{g}$ och $\mathbf{g}(\mathbf{t})$ är inre punkt till definitionsmängden till $\mathbf{f}$.

Komponenterna till sammansättningen av $\mathbf{f} \circ \mathbf{g}$ ges av:

$$(\mathbf{f} \circ \mathbf{g})_i(\mathbf{t}) = f_i(\mathbf{g}(\mathbf{t}))$$

Där f_i är en funktion $\mathbb{R}^p$ till $\mathbb{R}$ vilket låter oss använda den mindre generella kedjeregeln som är känd.

$$\frac{\partial}{\partial t_j}(\mathbf{f} \circ \mathbf{g})_i(\mathbf{t}) = \nabla f_i(\mathbf{g}(\mathbf{t}))^T \frac{\partial \mathbf{g}(\mathbf{t})}{\partial t_j} = \begin{bmatrix} \frac{\partial f_i}{\partial x_1} & \dots & \frac{\partial f_i}{\partial x_p} \end{bmatrix} \begin{bmatrix} \frac{\partial g_1}{\partial t_j} \\ \dots \\ \frac{\partial g_p}{\partial t_j} \end{bmatrix}$$

Detta ger rad i och kolonn j funktionsmatrisen till sammansättningen $(\mathbf{f} \circ \mathbf{g})'(\mathbf{t})$. Läggs nu det här ihop för i och j ger den allmänna formeln för kedjeregeln.

$$(\mathbf{f} \circ \mathbf{g})'(\mathbf{t}) = \mathbf{f}'(\mathbf{g}(\mathbf{t})) \mathbf{g}'(\mathbf{t})$$

Vilket är likt kedjeregeln för en variabel men skillnaden är att istället för funktioner och tal/variabler används vektorer av funktioner och vektorer av tal/variabler.

7 Implicita funktionssatsen

Inversa funktionssatsen

Låt $\mathbf{y} = \mathbf{f}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ vara $\mathcal{C}^1$. Låt även $\mathbf{a}$ vara en punkt i $\mathbf{f}$:s definitionsmängd, sådan att $\det(\mathbf{f}'(\mathbf{a})) \neq 0$. Då har $\mathbf{a}$ och $\mathbf{f}(\mathbf{a})$ varsin öppen omgivning U och V så att

$$\mathbf{f} : U \rightarrow V$$

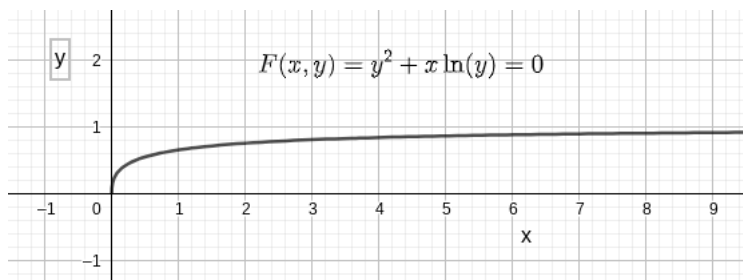
är bijektiv och dess invers,

$$\mathbf{f}^{-1} : V \rightarrow U$$

är $\mathcal{C}^1$.

OBS! $\det(\mathbf{f}'(\mathbf{a})) \neq 0 \ \forall \mathbf{a} \in \mathbb{R}^n$ medför ej att $\mathbf{f}$ är bijektiv i hela $\mathbb{R}^n$. Snarare medför det att varje $\mathbb{R}^n$ -punkt har en omgivning där bijektivitet råder.

Betrakta ekvationen $y^2 + x \ln(y) = 0$ då $x > 0$. Vi kan omöjligen lösa ut vad y är beroende på x , d.v.s. vi kan ej skriva $y = \text{känd funktion}(x)$, men y kan ändå ses som en funktion av x . Ty grafen nedan visar att för varje x -värde finns ett och endast ett y -värde. Det är detta som menas med en implicit funktion.



Figur 1:

Följande sats är användbar i sammanhang som dessa.

Implicita funktionssatsen

Låt $F(x, y)$ vara en $\mathcal{C}^1$ -funktion och (a, b) vara en punkt på nivåkurvan $F(x, y) = k$, $k \in \mathbb{R}$. Om

$$F'_y(a, b) \neq 0$$

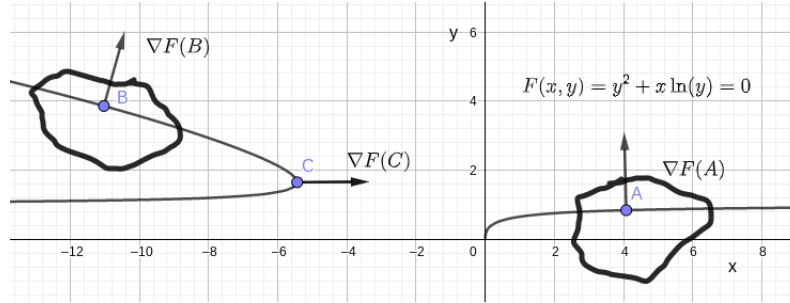
så finns en öppen omgivning U till (a, b) där inskränkningen av nivåkurvan till U implicit definierar en $\mathcal{C}^1$ -funktion $y = f(x)$. För denna gäller att

$$\frac{df}{dx} = -\frac{F'_x(x, f(x))}{F'_y(x, f(x))}.$$

Om vi återvänder till fallet ovan med $F(x, y) = y^2 + x \ln(y)$ och $k = 0$, men avskaffar $x > 0$, ser vi i bilden nedan att både punkt A och B har en omgivning (inom svart) där y är en funktion av x . Men i punkten C saknas sådan omgivning ty så fort vi rör oss det minsta åt vänster antar y två skilda värden, för varje x . Detta beror på att gradienten i C är vågrät. Att gradienten är parallell med x -axeln betyder att dess y -komponent är 0, d.v.s.

$$F'_y(C) = 0, \text{ i } \nabla F(C) = (F'_x(C), F'_y(C)).$$

Detta bryter mot antagandet i satsen.



Figur 2:

Implicita funktionssatsen gäller även om $F = F(y, x_1, x_2, \dots, x_n)$ och man vill se y som en funktion $y(x_1, \dots, x_n)$. På mest allmän form blir satsen som följande.

Låt $\mathbf{F} : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^n$ vara $\mathcal{C}^1$. Låt även $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^m$ uppfylla

$$\mathbf{F}(\mathbf{x}, \mathbf{t}) \Big|_{\mathbf{x}=\mathbf{a}, \mathbf{t}=\mathbf{b}} = \mathbf{C} \quad \mathbf{x} \in \mathbb{R}^n, \mathbf{t} \in \mathbb{R}^m$$

samt att $\frac{d(\mathbf{F})}{d(\mathbf{x})} \Big|_{(\mathbf{a}, \mathbf{b})} \neq \mathbf{0}$. Då definierar ekvationen $\mathbf{F}(\mathbf{x}, \mathbf{t}) = \mathbf{C}$ en funktion $\mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ som är $\mathcal{C}^1$, sådan att $\mathbf{x} = \mathbf{f}(\mathbf{t})$ i en omgivning av $(\mathbf{a}, \mathbf{b})$.